2005 Workplace and Equal Opportunity Survey of Active Duty Members:

Statistical Methodology Report

SURVEY OF ACTIVE DUTY MEMBERS:

STATISTICAL METHODOLOGY REPORT

DMDC

Defense Manpower Data Center
Survey & Program Evaluation Division
1600 Wilson Boulevard, Suite 400, Arlington, VA 22209-2593

Acknowledgments

SURVEY OF ACTIVE DUTY MEMBERS: STATISTICAL METHODOLOGY REPORT

Executive Summary

This report describes the sample design, sample selection, weighting, and variance estimation procedures for the *2005 Workplace and Equal Opportunity Survey of Active Duty Members* (*2005 WEOA*). The first part of this report presents the sample design and sample selection procedures. The second part provides information on the statistical methodology used for survey sample weighting. Response rates and location rates for the full sample and for subgroups are described in the final section of the second part.

The *2005 WEOA*, administered by the Department of Defense (DoD) Defense Manpower Data Center (DMDC) in 2005, collected information from active duty members with at least 6 months of service at the time of data collection. The *2005 WEOA* continued research in the area of workforce and gender relations that started with the DoD *1996 Equal Opportunity Survey (EOS)*.

The sample for the *2005 WEOA* consisted of a stratified random sample of 91,024 military members, of whom 32,299 were ultimately determined to be eligible members of the target population for the weighting process. The sample strata used for survey sample selection were Service (Army, Navy, Marine Corps, Air Force, and Coast Guard), race/ethnicity, paygroup, and region. The total sample size was based upon precision requirements for key reporting domains. A formal mathematical procedure was used to determine the sample allocation. The procedure involved developing equations to describe the variance of the sample estimates and the variable survey costs, then simultaneously solving the equations subject to precision requirements. The sample of individuals was selected with equal probabilities within strata; however, because the stratum allocations were not proportional to the stratum sizes, sampled members were not selected with equal overall probabilities.

The *2005 WEOA* weights were created in four steps. In the first step, an assigned final disposition code classified sampled members as eligible respondents, eligible nonrespondents, ineligible members, or members with unknown eligibility. The assignment of final disposition codes was a sequential process that drew upon sample selection, data collection, and returned questionnaire information. In the second step, a base weight, computed as the inverse of probability of selection, was assigned to each sample member. In the third step, base weights were adjusted for nonresponse in two stages. In the first stage, base weights were adjusted to account for members whose eligibility was not known at the end of data collection. In the second stage, base weights were adjusted to account for eligible members who returned incomplete or non-usable questionnaires. In the fourth and final step, the weights were raked to control totals to reduce bias not accounted for in the previous steps.

Since the *2005 WEOA* sample design was complex (not a simple random sample), specialized methods were required to account for the sample design during statistical processing. This issue is conveyed briefly in the main body of this report and is discussed more fully in Appendix D where the issue of variance estimation for complex surveys is discussed with

reference to linearization and replication strategies. Appendix F presents examples of processing the *2005 WEOA* survey using statistical software appropriate for complex surveys. Examples of both input code and procedural output are presented.

Response rates are generally used to measure the success and quality of survey administration. Survey location, completion, and response rates are reported in the second part of this report. More detailed information on response rates by population subgroups is contained in Appendix E. In reporting these rates, guidelines recommended by the Council of American Survey Research Organizations (CASRO) were followed. The weighted location, completion, and response rates for the *2005 WEOA* were 88.5%, 43.7%, and 38.7%, respectively.

Table of Contents

Table of Contents (Continued)

Appendixes

Table of Contents (Continued)

List of Tables

Table of Contents (Continued)

Table of Contents (Continued)

SAMPLE DESIGN FOR THE 2005 WORKPLACE AND EQUAL OPPORTUNITY SURVEY OF ACTIVE DUTY MEMBERS

Defense Manpower Data Center

Sampling Overview

The 2005 *Workplace and Equal Opportunity Survey of Active Duty Members* (*2005 WEOA*) was designed to represent active duty members of the Army, Navy, Marine Corps, Air Force, and Coast Guard, up to and including paygrade O-6, with at least six months service at the time the first questionnaire was mailed. A single-stage, stratified random sampling design was used to select members from the frame. The sampling frame consisted of 1,376,874 members from the Defense Manpower Data Center's (DMDC's) June 2004 Active Duty Master File (ADMF), Active Duty Pay File, Family Database File, Basic Allowance for Housing (BAH), and the Defense Enrollment Eligibility Reporting System (DEERS) File. The frame was stratified and, within each stratum, active duty members were sampled with equal conditional probabilities and without replacement. Stratum-level sample sizes were determined by variance constraints imposed on key parameter estimates for specified domains.

Inferential Requirements

The inferential requirements for a survey are described in terms of:

- a fully operational definition of the population of inferential interest (i.e., the target population);

- key parameters used in developing the design; and

- precision requirements for the survey, stated as functions of the maximum values of the variances to be associated with the sample estimates of the key parameters.

Population Definition

The population definition identifies all individuals for whom conclusions are to be reached or about whom inferences are to be made based on the survey data. The population of inferential interest for the *2005 WEOA* consisted of active duty members of the Army, Navy, Marine Corps, Air Force, and Coast Guard, up to and including paygrade O-6, with at least 6 months service at the time the first questionnaire was mailed. The sample for the *2005 WEOA* consisted of 91,024 individuals.

Key Reporting Domains

Key parameters used as the basis for the design are defined in terms of characteristics of the overall population, characteristics of subpopulations of special interest (key domains), tests of hypotheses (including standardized comparisons), and the relations that exist at populations levels among specified observation variables. For this survey, the key parameters were

prevalence rates, defined as the proportion of active duty personnel belonging to specified domains who report having various attitudes, opinions, and experiences as measured by the survey. Some of the factors used to define the key reporting domains are listed in Table 1. An initial set of candidate domains was generated by considering various combinations of, and crosses among, the factors. Because domain sizes interact with precision requirements, several iterations were required to develop domain definitions consistent with the objectives of the survey and the resources available to administer the survey.

Table 1.
Factors Defining Key Reporting Domains

Factor	Levels
Service (CSERVICE)	Army
	Navy
	Marine Corps
	Air Force
	Coast Guard
Region (EOSREGION)	U.S.
	Europe, Asia, Pacific Islands and Other
	Unknown
Race/Ethnicity (EOSRETH)	Non-Hispanic White
	Non-Hispanic Black
	Hispanic
	Native American
	Asian and Pacific Islander
	Other
	Unknown
Paygrade Group (EOSCPAY)	E1-E3
	E4, Unknown Enlisted
	E5-E6
	E7-E9
	W1 to O6, Unknown Officer
Gender (CSEX)	Male
	Female
	Unknown

Precision Requirements

In general, precision requirements are specified in terms of the maximum expected values of the sampling variances for key domain estimates. The sampling variances are functions of the sample size, sample distribution, population variances, and design prevalences.[1] A uniform prevalence rate of 0.50 was used to design the *2005 WEOA* sample.

For this survey, the maximum variances expected for particular sample results (estimates) were specified in terms of 95% confidence interval half-widths, or margins of error.[2] Both the cost implications and the objectives of the survey were considered in specifying these values. Appendix A, Table A-2, lists the half-width confidence interval set as precision requirements, together with domain definitions, and the estimated eligible population size for each domain.

Domains and their associated precision constraints were defined to allow separate in-depth analysis for each race/ethnic category in the overall active duty population, as well as for smaller domains also broken down by race/ethnic category. The survey precision requirements were set for domains to facilitate analyses both at the Armed Forces level and within the Services.

Sampling Frame Construction and Stratification

For sampling, a distinction was made between *dimensions of stratification* and *levels of stratification*. The dimensions are the variables used to stratify the sample/population whereas the levels are the values present within a dimension.

The following set of variable dimensions and levels were used to define strata for the sample:

- Service (CSERVICE): Army, Navy, Marine Corps, Air Force, Coast Guard;

- Paygrade group (EOSCPAY): E1-E3, E4, E5-E6, E7-E9, W1 to O6;

- Race/Ethnicity (EOSRETH): Non-Hispanic White, Non-Hispanic Black, Hispanic, Native American, Asian & Pacific Islander, Other; and

- Region (EOSREGION): U.S. or Not U.S. (Europe, Asia, Pacific Islands, Other).

Stratification

As a starting point, a candidate set of strata was constructed by crossing all the levels of the stratification variables, yielding 300 potential strata. The next step was to consider minimum stratum size consistent with a total sample size of 91,024. If unbiased variances for linear statistics are to be a design requirement, then a minimum of two observations is needed in any stratum. However, if a stratum is too small, then insisting on two observations from that stratum

---

[1] Prevalence rates are the proportion of persons belonging to specified domains who would report having the various attitudes and experiences measured on the survey.

[2] *Margins of error*, such as those reported for opinion polls, are expressed as plus or minus figures. The *confidence level*, typically 95%, represents the probability that the true population value is covered by the confidence interval in repeated samples.

introduces an unequal weighting effect that acts to increase variances for no reason other than the stratum is too small. Even if only a few strata are too small, the cumulative unequal weighting effects can compromise any variance advantage associated with having stratified in the first place.

This consideration led to defining "too small" in terms of a proportional allocation of the total sample. A proportional allocation of the sample cannot, by definition, introduce unequal weighting effects. Given a proportional allocation and a minimum requirement of two observations per stratum, the minimum stratum size was computed as,

$$\min\{N_h\} = \frac{2N}{n}$$

where,

N_h = the size of the hth stratum;

N = the size of the population; and

n = the total size of the sample.

For N = 1,376,874 and n = 91,024, a minimum stratum size of $\min\{N_h\}$ = 30.3 (rounded to 50) was adopted.

Next, the proportion of the total strata defined by all possible crosses that were below the minimum size of 30 was computed for each of the initial stratification variables. The decisions about which strata to collapse were based on identifying the candidate stratification dimensions with consistent patterns of deficient strata and on consideration of the relative importance of specific candidate stratification dimensions to the surveys. Specific levels that were collapsed were:

- U.S. and Not U.S.;

- Native American and Other; and

- Paygrade Groups E5-E6 and E7-E9.

The final strata definitions are listed in Appendix A, Table A-1. A total of 220 strata were constructed. The "unknown" stratum (stratum 220 in Table A-1) contains members for whom one or more of the stratum dimensions were missing in the sampling frame.

Sample Size Allocation

After the *2005 WEOA* strata were constructed, domains and their associated precision constraints were defined. Precision requirements were set for selected domains to allow in-depth analysis for the overall active duty population and some depth of analysis for other domains. Special attention was given to allow for Service-level analyses.

After the strata were constructed, the total sample size and its allocation to the sampling strata were determined. The DMDC sampling tool (Kavee and Mason, 1997) was used to allocate the sample so that the precision requirements were met, in expectation, for the different reporting domains. This software is designed to produce optimal sample designs for stratified, equal probability within-stratum samples for a specified cost model. The cost model used is described by Wheeless, Mason, and Kavee (1997).

WEIGHTING FOR THE 2005 WORKPLACE AND EQUAL OPPORTUNITY SURVEY OF ACTIVE DUTY MEMBERS

Westat

Weighting Overview

This chapter describes weighting procedures implemented for the Department of Defense's (DoD) *2005 Workplace and Equal Opportunity Survey of Active Duty Members* (*2005 WEOA*). The *2005 WEOA*, administered by the DoD Defense Manpower Data Center (DMDC), was sponsored by the Office of the Under Secretary of Defense (Personnel and Readiness). The *2005 WEOA* was the second equal opportunity survey of active duty members conducted by DMDC. The first survey was the *Armed Forces 1996 Equal Opportunity Survey (1996 EOS)* (Mason, Kavee, Wheeless, George, Riemer, and Elig, 1996). Differences in weighting methodology and assignment of final disposition codes between the *2005 WEOA* and the *1996 EOS* are discussed, as well as how comparisons of estimates between these surveys can be made.

The *2005 WEOA* analytical weights were created in four steps. In the first step, sampled members were classified using an assigned final disposition code as eligible respondents, eligible nonrespondents, ineligible members, or members with unknown eligibility. The assignment of final disposition codes was a sequential process that drew upon information from the updated sampling frame, field operations, and returned questionnaire information. In the second step, a base weight, computed as the inverse of probability of selection, was assigned to each sample member. In the third step, base weights were adjusted for nonresponse in two stages. In the first stage, base weights were adjusted to account for members whose eligibility was not known at the end of data collection. In the second stage, the weights were adjusted to account for eligible members who returned incomplete or unusable questionnaires (Appendix B details nonresponse adjustments). In the fourth and last step, the weights were raked to frame control totals to reduce bias not accounted for in the previous steps. This final adjustment compensated for changes in the population that occurred between the time of sample selection and data collection. (Appendix C presents the dimensions used during raking.)

Since the *2005 WEOA* sample design was complex (not a simple random sample), specialized methods were required to account for sample design during statistical processing. Appendix D presents approaches for variance estimation for complex surveys.

Response rates for the *2005 WEOA* have been computed in accordance with the standards defined by the Council of American Survey Research Organizations (CASRO, 1982). The response rates for the full sample and subgroups, and how they were computed, are described in the last section of the body of this report. Appendix E details these calculations.

Finally, methods of computing variance estimates for the *2005 WEOA* and comparing and combining *2005 WEOA* survey data with other surveys are discussed with software input code and output examples in Appendix F.

Eligibility in the 2005 WEOA

As in most surveys, there was a lag between sample selection and data collection. During this lag there was attrition in the target population. Some members separated, retired, were promoted to ineligible paygrades,[3] or died between the creation of the sampling frame (June 2004) and the beginning of data collection (January 2005). In other words, some members changed their survey eligibility status *after* the sampling frame was created. Part of this attrition was identified prior to data collection through the use of more recent administrative files. Other attrition was identified during survey administration. However, information could not be determined on the attrition of members who either did not receive a questionnaire (because of bad or incomplete mailing addresses) or did not return the survey.

Analytical weights were created so that estimates from the survey represent the population of interest. These weights reflect the probability of selection and nonresponse adjustment factors computed to minimize bias due to differential response rates among demographic subgroups of the population. During weighting, the weights of respondents were adjusted to represent nonrespondent members, but weights for ineligible sample members were not generally adjusted in the same way as those for nonrespondents. Therefore, it was critical to determine which nonresponding sample members should be coded as ineligible before the weights were adjusted for nonresponse. In the *2005 WEOA*, active duty members of the Army, Navy, Marine Corps, Air Force, and Coast Guard, up to and including paygrade O-6, were eligible for the survey if the member met the following conditions:

- Member had at least six months of service at the time the first questionnaire was mailed;

- Member was eligible in the June 2004 Active Duty Master File (ADMF) sampling frame, the January 2005 ADMF updated frame, and the February 2005 Defense Enrollment Eligibility Reporting System (DEERS); and

- Member self-reported (or by proxy) that he was on active duty on January 24, 2005 (i.e., not retired or separated).

This eligibility definition is consistent with prior DMDC surveys and other similar surveys where the eligibility of the sampled units changes over time. This definition of ineligibility recognizes that there is attrition in the member population due to promotion, separation, retirement, hospitalization, death, or incarceration. Consequently, the sum of the analytical weights (adjusted for nonresponse) is an estimate of the surviving population at the time of data collection.

---

[3] Members who were promoted to paygrade O7 or above were not eligible for the survey. These members were excluded during the creation of the updated frame.

Assigning Disposition Codes for the 2005 WEOA

Each sampled member in the *2005 WEOA* was assigned a disposition code (variable ELIG_R) with the member's response disposition code for the survey. These codes were a key input in weighting and in the computation of response rates. The procedure for deriving the value of ELIG_R for each sampled person involves several steps that are described in the following sections. The response disposition code included the following groups or categories:

- *ER*—Eligible respondents: this group consists of all eligible members who participated in the survey and provided substantially complete and usable survey data;

- *ENR*—Eligible nonrespondents: this group consists of all sampled members who were known, or assumed, to be eligible for the survey, but did not provide substantially complete and usable survey data;

- *IN_FR*—Frame ineligible members or out-of-scope members as determined by the updated January 2005 frame file: this group consists of all sampled members determined to be ineligible prior to the beginning of data collection;

- *IN_PR*—Self-reported or proxy-reported ineligible members: this group includes those members who self-reported or were reported by proxy as being separated, retired, deceased or incarcerated; and

- *UNK*—Other nonrespondent members whose eligibility is unknown: this group consists of all the members whose eligibility could not be determined (for example, postal non-deliverables, other non-locatables, and members who did not return the questionnaire).

The assignment of disposition codes drew upon information from a number of sources. The assignment was a sequential process that used the following variables created during sample selection and data collection:

- Variable F_ELIG — Updated January 2005 frame eligibility indicator;

- Variable RFLAG_FIN—Survey Control System disposition code;

- Variable SCSINEL—Reason for reported ineligibility from the Survey Control System.

- Variable SR_ELIG—Self-reported eligibility; and

- Variable COMPFLAG—Completed questionnaire indicator.

The creation and description of these variables are presented in the following subsections.

Frame Eligibility

Westat created the variable F_ELIG to indicate the frame eligibility of the member as of January 2005 (beginning of the data collection period). This variable reflects the eligibility of the member using information from the January 2005 ADMF file. The variable F_ELIG was assigned for all the records in the June 2004 sampling frame, using the variables INJUN (in June 2004 frame indicator) and INJAN (in January 2005 frame indicator), which were created by merging the June 2004 sampling frame with the updated January 2005 frame. DMDC provided a file with the January frame restricted to members in the June frame who were still eligible in January 2005. Table 2 shows how the variable F_ELIG was created. A member was frame eligible (F_ELIG = 1) if the member was eligible in the June sampling frame (INJUN = 1) and eligible in the January frame (INJAN = 1). After merging the files, 57,466 members (4.17 %) were classified as frame ineligible (F_ELIG = 0). The control totals used to benchmark the final weights were derived using all records in the frame with F_ELIG = 1.

Table 2.
Distribution of 2005 WEOA Frame Eligibility (Variable F_ELIG) in the Population

F_ELIG	INJUN	INJAN	Frame Cases	Percentage of Cases in the Frame
1 – Frame eligible member	1	1	1,319,408	95.83
0 – Frame ineligible member	1	2	57,466	4.17
Total			**1,376,874**	**100.00**

Table 3 summarizes the distribution of the variable F_ELIG in the sample. Because all ineligible members of this type were identified in the sample, these members were assigned a specific disposition code (*IN_FR*), to distinguish them from ineligible members identified during data collection (*IN_PR*). As indicated in the table, there are 3,609 sampled members (3.96 % of the sample) classified as frame ineligible (F_ELIG = 0).

Table 3.
Distribution of 2005 WEOA Frame Eligibility (Variable F_ELIG) of Sampled Members at the Time of Data Collection

F_ELIG	INJUN	INJAN	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage Sum of Base Weights
1 – Eligible	1	1	87,415	96.04	1,319,994	95.87
0 – Ineligible	1	2	3,609	3.96	56,880	4.13
Total			**91,024**	**100.00**	**1,376,874**	**100.00**

In previous DMDC surveys, the information from DEERS was used to identify additional frame ineligible members during the creation of the variable F_ELIG. In the *2005 WEOA*, the information from DEERS was only used identify ineligible members in the sample before mailing out the questionnaires. An analysis of the estimate of DEERS ineligible members shows only a very small impact on the number of frame ineligible members. An additional 0.05% of members in the sample and frame could have been identified as frame ineligibles (F_ELIG=0) if the information in DEERS had been used in the creation of F_ELIG.

Survey Control System Disposition

The Survey Control System (SCS) used for survey operations contained the variable FLAG_FIN with the field operation disposition code of each mailed survey. The variable RFLAG_FIN[4] was created during data collection and the values were assigned based on the results of the mailing waves (e.g., sent/received questionnaires, postal non-deliverables, non-locatable) and condition of the returned questionnaire (blank/non-blank). Table 4 shows the sample distribution and descriptions of the levels of the variable RFLAG_FIN.

Table 4.
Description and Distribution of the Survey Control System Disposition Codes (Variable RFLAG_FIN)

RFLAG_FIN	Description	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
1	Returned survey - a non-blank survey was returned with no additional information.	35,568	39.08	557,481	40.49
2	Return (deceased) – a non-blank survey was returned with additional information that the sample member was deceased.	4	0.00	27	0.00
6	Return – a non-blank survey was returned with additional information that the sample member was separated/retired.	15	0.02	272	0.02
7	Return – a non-blank survey was returned with additional information that the sample member was deployed.	85	0.09	1,350	0.10

---

[4] RFLAG_FIN was created by recoding the SCS variable FLAG_FIN using the variable ELIG0410. The variable ELIG0410 is a mailing eligibility flag created by DMDC using the DEERS file. The variable FLAG_FIN misclassified the original ineligible members identified by DMDC (FLAG_FIN=30). The variable RFLAG_FIN fixed this problem.

Table 4.   (Continued)

RFLAG_FIN	Description	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
8	Return (all other reasons) – a non-blank survey was returned with a reason other than that the sample member was deceased, or incarcerated.	94	0.10	1,350	0.10
13	Returned Blank – a blank survey was returned with information that the sample member was separated/retired.	2	0.00	50	0.00
14	Returned Blank (active refusal) – a blank survey was returned, sample member refused to take part in the survey.	5	0.01	92	0.01
15	Returned Blank – a blank survey was returned with information that the sample member was deployed.	22	0.02	346	0.03
17	Returned Blank (no reason) – a blank survey was returned and no reason was given by sample member.	3	0.00	45	0.00
18	No Return (deceased) – survey was not returned, sample member deceased.	33	0.04	448	0.03
19	No Return (incarcerated) – survey was not returned, sample member was incarcerated.	3	0.00	70	0.01
22	No Return – survey was not returned, sample member was separated/retired.	132	0.15	2,012	0.15
23	No Return (active refusal) – survey was not returned, or sample member refused to take part in the survey but did not identify himself as incarcerated.	91	0.10	1,188	0.09
24	No Return (deployed) – survey was not returned, sample member unreachable at UNIT address because of deployment	405	0.44	5,521	0.40
25	No Return (all other reasons) – survey was not returned, sample member did not actively refuse, gave a reason for nonresponse other than being deceased, incarcerated.	10	0.01	168	0.01

Table 4.   (Continued)

26	No Return (no reason) – survey was not returned, no reason was given by sample member	41,776	45.90	622,591	45.22
27	PND (no address remaining) – all addresses were attempted-returned PND	3,195	3.51	46,648	3.39
28	PND (address remaining) – addresses were attempted-returned PND with addresses remaining at close of field	8,370	9.20	117,659	8.55
29	Original Non-Locatable (no address at start of mailing) – substantially incomplete or blank address field prior to the start of the administration of the survey, no mailings attempted	3	0.00	48	0.00
30	Original ineligible as identified by DMDC	1,208	1.33	19,509	1.42
Total		91,024	100.00	1,376,875	100.00

Reason Reported for Ineligibility Variable

The Survey Control System (SCS) contained the variable SCSINEL, which is the reason reported for ineligibility. This variable referred to member ineligibility from the point of view of field operations and did not necessarily match the member ineligibility used in weighting. A SCSINEL value of 8 was used in conjunction with RFLAG_FIN values of 8 and 25 to identify members considered ineligible for weighting due to illness. Members were assigned the RFLAG_FIN value of 8 if they returned a non-blank survey with a reason other than that the member was deceased or incarcerated. The RFLAG_FIN value of 25 denoted members who did not return a survey, but did not actively refuse, and who gave a reason for nonresponse other than being deceased or incarcerated. All other members with values of 8 or 25 on RFLAG_FIN who did not report being ill (SCSINEL $\neq$ 8) were considered eligible. Members with values of SCSINEL that indicated the member was deceased (2), incarcerated (7), separated (9), or retired (12) were identified as ineligible through other values of RFLAG_FIN. Being deployed or "Other" (SCSINEL = 13 or 14), was not a reason for ineligibility. The variable SCSINEL incorrectly coded these cases as ineligible. However, these cases were corrected using the variable RFLAG_FIN. Table 5 shows the sample distribution of the variable SCSINEL.

Table 5.
Distribution of the Member's Reason for Ineligibility (Variable SCSINEL)

SCSINEL	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
0 – Not ineligible	90,296	99.20	1,366,394	99.24
2 – Deceased	53	0.06	693	0.05
7 – Incarcerated	3	0.00	70	0.01
8 – Ill	9	0.01	127	0.01
9 – Separated	62	0.07	964	0.07
12 – Retired	88	0.10	1,390	0.10
13 – Other	4	0.00	70	0.01
14 – Deployed	509	0.56	7,165	0.52
Total	91,024	100.00	1,376,874	100.00

To facilitate the creation of the disposition code for the *2005 WEOA*, the variable RSCSINEL was created by recoding the variables SCSINEL, as shown in Table 6. This table shows the distribution of the variable RSCSINEL in the sample.

Table 6.
Distribution of the Variable RSCSINEL (Reason Reported for Ineligibility)

RSCSINEL	SCSINEL	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
O – Other	0, 2, 7, 9, 12, 13, 14	91,015	99.99	1,376,747	99.99
I– Ill	8	9	0.01	127	0.01
Total		91,024	100.00	1,376,874	100.00

The variable RSCSINEL had not been used in DMDC surveys before 2004, instead, all members with RFLAG_FIN values of 8 and 25 (reasons other than deceased, incarcerated, separated, retired or deployed) were considered eligible members.

Self-Reported Eligibility

The sampled members were asked in which Service branch they were on active duty as of January 24, 2005 (Question 2, variable SRSVC1). Table 7 shows the distribution of the variable SR_ELIG in the sample.

Table 7.
Distribution of Self-reported Service as of January 24, 2005 (Variable SRSVC1)

SRSVC1	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
No response	166	0.18	2,715	0.20
No survey return	55,226	60.67	815,861	59.25
Not applicable	921	1.01	12,987	0.94
1 – Army	12,724	13.98	155,282	11.28
2 – Navy	8,101	8.90	143,802	10.44
3 – Marine Corps	4,463	4.90	38,744	2.81
4 – Air Force	7,796	8.56	185,193	13.45
5 – Coast Guard	1,400	1.54	18,260	1.33
6 – None, you were separated or retired	227	0.25	4,030	0.29
Total	91,024	100.00	1,376,874	100.00

The variable SR_ELIG was created to indicate the self-reported eligibility of the members. Respondents who indicated that they were either separated or retired as of that date were considered ineligible. Table 8 shows the distribution and the creation of the variable SR_ELIG.

Table 8.
Distribution of Self-reported Eligibility (Variable SR_ELIG)

SR_ELIG	SRSVC1	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
1 – Self reported Eligible	1, 2, 3, 4, 5, missing	90,797	99.75	1,372,844	99.71
2 – Self reported Ineligible	6	227	0.25	4,030	0.29
Total		91,024	100.00	1,376,874	100.00

Completed Questionnaire Indicator

DMDC created the variable COMPFLAG that indicates whether a questionnaire was completed. A questionnaire was considered complete if more than 50% of a required group of questions were answered by the sampled member. The sample distribution of the variable COMPFLAG is shown in Table 9.

Table 9.
Distribution of Completed Questionnaires (Variable COMPFLAG)

COMPFLAG	Sample Cases	Percentage of Sample Cases	Sum of Base Weights	Percentage of Sum of Base Weights
1 – Complete	32,401	35.60	510,340	37.07
0 – Incomplete	3,397	3.73	50,673	3.68
.B – No survey return	55,226	60.67	815,861	59.25
Total	91,024	100.00	1,376,874	100.00

Final Disposition Codes

The method used to assign final disposition codes (variable ELIG_R) was a sequential process using the variables described above. Figure 1 shows the process of assigning values of the variable ELIG_R.

The main rule was to identify first any ineligible members using the available data sources (i.e., updated frame, self-reported eligibility, and SCS information). First, sampled members found to be ineligible before data collection (variable F_ELIG = 0) were coded as frame ineligible (ELIG_R = *IN_FR*). Then, the values of the variable SR_ELIG were examined. Members who reported being separated or retired were coded as self-reported ineligible (*IN_PR*). Next, the values of the variable RFLAG_FIN were inspected. Members who did not return the questionnaire were classified as eligible nonrespondents (*ENR*), proxy ineligible (*IN_PR*), or eligibility unknown (*UNK*) depending on the value of RFLAG_FIN, and RSCSINEL. Members who returned the questionnaire with the variable RFLAG_FIN = 1 or 7 were classified as eligible respondents (*ER*) or eligible nonrespondents (*ENR*) based on whether the questionnaire was complete, as determined by the variable COMPFLAG.

After assigning disposition codes, all combinations of variables used to create ELIG_R were checked for inconsistencies. All inconsistencies were reported to DMDC for review. Table 10 lists the various combinations of the variables ELIG_R, F_ELIG, RFLAG_FIN, SR_ELIG, RSCSINEL, and COMPFLAG that occurred in the *2005 WEOA*, along with the numbers of sampled cases in each combination and the sum of base weights.

Figure 1.
Sequential Assignment of ELIG_R Disposition Codes

START → Check F_ELIG

F_ELIG=1 —NO→ Code as IN_FR

YES ↓

CHECK SR_ELIG

SR_ELIG=1 —NO→ Code as IN_PR

YES ↓

CHECK RFLAG_FIN

RFLAG_FIN=1, 7 —YES→ (to: Is questionnaire complete?)

NO ↓

RFLAG_FIN=26,27,28,29 —YES→ Code as UNK

NO ↓

RFLAG_FIN=8,25 —YES→ Check RSCSINEL   RSCSINEL = I —NO→ (to: Is questionnaire complete?)

YES ↓ (RSCSINEL = I) → Code as IN_PR

NO ↓

RFLAG_FIN=2,3,6,9,10,13, 18,19,22,30 —YES→ Code as IN_PR

NO ↓

RFLAG_FIN=14,15,17,23,24 —YES→ Code as ENR

NO ↓

Is questionnaire complete?   COMPFLAG = 1 —NO→ Code as ENR

YES → Code as ER

16

Table 10.

Combinations of Variables Used to Determine Disposition Codes for the 2005 WEOA (Variable ELIG_R)

Row	Eligibility (ELIG_R)	Frame Eligibility (F_ELIG)	Survey Control System Disposition Code (RFLAG_FIN)	SCS Eligibility (RSCSINEL)	Self-reported Eligibility (SR_ELIG)	Complete Questionnaire (COMPFLAG)	N	Sum of Base Weights
Eligible Respondents								
1	ER	1	001 - Returned survey	O	1	1	32,165	506,836
2	ER	1	007 - Return (deployed)	O	1	1	65	989
3	ER	1	008 - Return (all other reasons)	O	1	1	69	993
Total ER							**32,299[a]**	**508,818**
Eligible Nonrespondents								
4	ENR	1	001 - Returned survey	O	1	0	3,058	44,874
6	ENR	1	007 - Return (deployed)	O	1	0	20	361
7	ENR	1	008 - Return (all other reasons)	O	1	0	23	335
8	ENR	1	014 - Returned Blank (active refusal)	O	1	0	3	66
9	ENR	1	015 - Returned Blank (deployed)	O	1	0	22	346
10	ENR	1	017 - Returned Blank (no reason)	O	1	0	2	44
11	ENR	1	023 - No Return (active refusal)	O	1	.B	90	1,170
12	ENR	1	024 - No Return (deployed)	O	1	.B	401	5,466
13	ENR	1	025 - No Return (all other reasons)	O	1	.B	3	64
Total ENR							**3,622**	**52,726**
Ineligible as Self-reported or Reported by Proxy								
5	IN_PR	1	001 - Returned survey	O	2	0	39	788
18	IN_PR	1	008 - Return (all other reasons)	I	2	0	1	5
17	IN_PR	1	008 - Return (all other reasons)	I	1	1	1	18
23	IN_PR	1	025 - No Return (all other reasons)	I	1	.B	7	104
14	IN_PR	1	002 - Return (deceased)	O	1	1	4	27
15	IN_PR	1	006 - Return (separated/retired)	O	1	0	1	22
16	IN_PR	1	006 - Return (separated/retired)	O	1	1	7	130
19	IN_PR	1	013 - Returned Blank (separated/retired)	O	1	0	1	5
20	IN_PR	1	018 - No Return (deceased)	O	1	.B	6	61
21	IN_PR	1	019 - No Return (incarcerated)	O	1	.B	3	70
22	IN_PR	1	022 - No Return (separated/retired)	O	1	.B	43	752
24	IN_PR	1	030 - Original ineligible as identified by DMDC	O	1	.B	76	716
Total IN_PR							**189**	**2,698**

Table 10.   (Continued)

Row	Eligibility (ELIG_R)	Frame Eligibility (F_ELIG)	Survey Control System Disposition Code (RFLAG_FIN)	SCS Eligibility (RSCSINEL)	Self-reported Eligibility (SR_ELIG)	Complete Questionnaire (COMPFLAG)	N	Sum of Base Weights
Ineligible as Reported by the Frame								
26	IN_FR	2	001 - Returned survey	O	1	0	33	477
27	IN_FR	2	001 - Returned survey	O	1	1	89	1,297
28	IN_FR	2	001 - Returned survey	O	2	0	184	3,209
29	IN_FR	2	006 - Return (separated/retired)	O	1	0	3	41
30	IN_FR	2	006 - Return (separated/retired)	O	1	1	1	50
31	IN_FR	2	006 - Return (separated/retired)	O	2	0	3	28
32	IN_FR	2	013 - Returned Blank (separated/retired)	O	1	0	1	45
33	IN_FR	2	014 - Returned Blank (active refusal)	O	1	0	2	26
34	IN_FR	2	017 - Returned Blank (no reason)	O	1	0	1	1
35	IN_FR	2	018 - No Return (deceased)	O	1	.B	27	387
36	IN_FR	2	022 - No Return (separated/retired)	O	1	.B	89	1,260
37	IN_FR	2	023 - No Return (active refusal)	O	1	.B	1	18
38	IN_FR	2	024 - No Return (deployed)	O	1	.B	4	55
39	IN_FR	2	026 - No Return (no reason)	O	1	.B	1,181	19,048
40	IN_FR	2	027 - PND (no address remaining)	O	1	.B	315	4,467
41	IN_FR	2	028 - PND (address remaining at the close of field)	O	1	.B	543	7,678
42	IN_FR	2	030 - Original ineligible as identified by DMDC	O	1	.B	1,132	18,794
Total IN_FR							**3,609**	**56,881**
Unknowns								
45	UNK	1	026 - No Return (no reason)	O	1	.B	40,595	603,543
46	UNK	1	027 - PND (no address remaining)	O	1	.B	2,880	42,181
47	UNK	1	028 - PND (address remaining at the close of field)	O	1	.B	7,827	109,980
48	UNK	1	029 - Original Non-Locatable	O	1	.B	3	48
Total UNK							**51,305**	**755,752**
Grand Total							**91,024**	**1,376,875**

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes.
  This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

The information from the variables SR_ELIG, and RSCSINEL identified 48 sample members out of 189 previously ineligible members (i.e., coded as *IN_PR*). These records represented 33.9% of the estimated total ineligible members coded as *IN_PR* in the population, and an estimated 0.07% of the total population in the frame. If the variables RSCSINEL and SR_ELIG had not been used, the estimate of ER members would have been 508,836 instead of 508,818; the estimate of *ENR* members would have been 53,623 instead of 52,726; and the estimate of *IN_PR* members would have been 1,783 instead of 2,698. These results suggested that, although the variable SCSINEL and SR_ELIG were used to identify additional *IN_PR* ineligible members in the sample, the impact on the total of ineligible members coded as *IN_PR* was negligible in the 2005 WEOA.

When assigning the value for ELIG_R, members who returned the questionnaire were assumed to be *eligible* (*ER*) unless they indicated otherwise. In particular, members with values of FLAG_FIN = 14, 15, 16, 17, 23 and 24 were coded as *eligible nonrespondents* (*ENR*). This group included all blank and non-blank returned questionnaires with reasons such as active refusal, member deployed, no reason, and all other reasons, except when the member was separated, hospitalized, deceased, retired, or incarcerated. This assumption is consistent with the assignment of disposition codes in the *1999 Survey of Active Duty Personnel* (*ADS*) (Wright, Elig, Flores Cervantes, George, & Valliant, 2000) *(Form B)*, the *2000 Reserve Component Survey* (*2000 RCS*) (Elig, Riemer, Simmons, & Valliant, 2002) *(Forms M* and *S),* the *2002 Workplace and Gender Relations Survey* (*2002 WGR*) (George & Kroeger, 2002*),* the *2003 Survey of Retired Military* (*2003 SRM*) (DMDC, 2004), and the *2004 Workplace and Gender Relation Survey of Reserve Component Members* (Riemer, 2004). This is different from the disposition code assignment in the *1999 ADS (Form A)*, where such cases were coded as members with *unknown eligibility* (*UNK)*; that is, members were eligible only if they explicitly indicated they were eligible.

The value of ELIG_R = *IN_PR* was assigned in a way similar to the *1999 ADS*, where a survey question asked whether a member was still in the armed forces. This self-report was used, in addition to FLAG_FIN codes on the SCS, to assign values *IN_PR*[5]. The assignment was somewhat different than in some other DMDC surveys such as the *1996 Equal Opportunity Survey (1996 EOS)* where the survey did not include a question to determine eligibility.

The method for assigning final disposition codes in some prior DMDC surveys differed from the method used in the 2004 WGRR. The central difference is in the treatment of eligible nonrespondents (*ENR*) and ineligible members identified though administrative files (*IN_FR*). In surveys such as the 1995 Sexual Harassment Survey (Mason, Kavee, Wheeless, George, Riemer, & Elig, 1996), the 1996 Armed Forces Equal Opportunity Survey (Wheeless, Mason, & Kavee, 1997) and the 1996 Retired Military Personnel Survey (Riemer & Lamoreaux, 2002), the *ENR* and *IN_FR* cases represented a relatively small number of sample members. Because of the small numbers, the *ENR*

---

[5] The value of ELIG_R=IN_SR in the *1999 ADS* is equivalent to the value of ELIG_R=IN_PR in the *2005 WEOA*.

members identified as eligible, but who returned incomplete and unusable surveys, were treated as members with unknown eligibility (*UNK*). In addition, the *IN_FR* members were treated as self- or proxy reported members (*IN_PR*) and were adjusted for nonresponse. Because there was only one group of nonrespondents (i.e., *UNK*) in these surveys, the weights of eligible respondents (*ER*) and ineligible members (*IN_PR* and *IN_FR*) were adjusted for nonresponse in one single step.

Refinements in the weighting process, implemented following the surveys identified above, required changes in the definition and assignment of final disposition codes. These changes were first introduced in the *1999 Active Duty Personnel Surveys* and continued in the *2000 Surveys of Reserve Component Personnel*, *2002 Workplace and Gender Relations Survey*, *2003 Survey of Retired Military*, and the *2004 Workplace and Gender Relations Survey of Reserve Component Members*. The changes required the creation of separate disposition codes for eligible nonrespondents (*ENR*) and ineligible members identified though the administrative files (*IN_FR*). These groups were handled differently through the weighting process as described in the following section. The creation of these groups provided a more precise adjustment for ineligible members in the sampling frame, and was more consistent with the response groups used to compute response rates in accordance with the standards defined by the Council of American Survey Research Organizations (CASRO).

Table 11 shows the distribution (number of sample cases and sums of base weights) of the final *2005 WEOA* disposition codes. Slightly more than 35% of the sample cases were coded as eligible respondents. Since most ineligible members were excluded from the updated frame, the remaining percentage of sample members coded as ineligible, as reported by self or proxy, was very small (0.20%).

Table 11.
Distribution of Final Disposition Code (Variable ELIG_R)

Variable ELIG_R		Sample Cases	Percent	Sum of Base Weights	Percentage of Sum of Base Weights
ER	Eligible respondents	32,299[a]	35.48	508,818	36.95
ENR	Eligible nonrespondents	3,622	3.98	52,725	3.83
IN_FR	Ineligible members as determined by the updated January 2005 frame and mail eligibility (DEERS)	3,609	3.96	56,880	4.13
IN_PR	Self- and proxy-reported ineligible members	189	0.21	2,700	0.20
UNK	Members with unknown eligibility	51,305	56.36	755,751	54.89
Total		**91,024**	**100.00**	**1,376,874**	**100.00**

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

Weighting Procedures

The analysis of survey data from complex sample designs requires the use of weights to (1) compensate for variable probabilities of selection; (2) adjust for differential response rates; and (3) improve the precision of the survey-based estimates (Skinner, Holt, and Smith (Eds.), 1989). To develop the analytical weights for the *2005 WEOA*, the following steps were taken. First, base weights, equal to the reciprocal of the probability of selection, were assigned to each sampled member. Next, the base weights were adjusted for nonresponse using weighting classes defined by the sampling stratum and relevant variables available from administrative record files compiled during the creation of the *2005 WEOA* sampling frame. In the last step, the nonresponse-adjusted weights were raked (ratio-adjusted) to population totals computed from the sampling frame. The raking adjustment compensated for any residual biases not accounted for by the nonresponse adjustments. Details of the weighting methodology are described below.

Calculation of Base Weights

The *2005 WEOA* sample was randomly selected without replacement from a stratified frame. As such, the sample size and overall probabilities of selection varied by sampling strata. The sampling strata and sample sizes were developed to satisfy the precision goals for domains of interest specified in the study. Let U be the frame of the N units in the population (i.e., active duty members at the time of sampling). Note that the frame includes some units who were ineligible at the time the survey was administered because, for example, they had died in the interval between sample selection and survey administration. The frame U was partitioned into H non-overlapping strata, $U_1, \ldots, U_H$, consisting of N_h units in each stratum h so that

$$N = \sum_{h=1}^{H} N_h.$$

A simple random sample of size n_h was selected without replacement within each stratum U_h. Given this design, the base weight for the ith sampled member in stratum h was calculated as:

$$w_{hi} = \frac{N_h}{n_h} \qquad i = 1, \ldots, n_h.$$

For each member classified in stratum h, the base weight was computed as the ratio of the total number of members in the stratum to the stratum-level sample size. The base weight w_{hi} is equal to the reciprocal of the probability of selection and was attached to each sample unit in the data file. Note that n_h is the number of members initially sampled in stratum h without regard to whether the member participated in the survey.

Nonresponse Weighting Adjustments

In an ideal world, all members of the inference population are eligible to be selected into the sample and all who are selected participate in the survey. In practice, neither of these conditions usually occurs. Some of the sampled members do not respond (survey nonresponse); some sample members are discovered to be ineligible (e.g., die or separate); and the eligibility status of some members cannot be determined. If these problems are not addressed, survey estimates will be biased. Special weighting adjustments were applied to the base weights to reduce the bias caused by nonresponse and unknown eligibility. Nonresponse-adjusted weights were created by multiplying the base weights by nonresponse adjustment factors. The following sections describe the statistical theory behind the nonresponse adjustments, including a description of the method used in the *2005 WEOA* to adjust the weights.

Unit Nonresponse Adjustments

Unit nonresponse occurs when a sampled member fails to respond for any reason. For example, unit nonresponse could result from failure to locate the member because of mobility, invalid/incorrect addresses in the frame, or from the unwillingness of some members to participate in the survey. Because the overall unweighted response rate in the survey was 37.14%, adjusting for unit nonresponse was an important step in attempting to minimize bias.

A potential drawback to nonresponse adjustments is that they can increase the variability of the weights and, thus, increase the sampling variance of some estimates (Kish, 1965). Response adjustments are beneficial only when the reduction in bias more than compensates for the increase in variance. Depending on the specific method used to adjust the weights, the sizes of the nonresponse adjustment factors are often constrained so they do not become either inordinately large or substantially different from each other. In most cases, the effect of the adjustments is modest.

This method used to adjust the base weights for nonresponse in the *2005 WEOA* is referred to as sample weighting or weighting class adjustment (Brick & Kalton, 1996). In this method, nonresponse adjustments are computed and applied separately by cells or weighting classes. A weighting class is created using characteristics known for both respondents and nonrespondents. Nonrespondents are assumed to be randomly distributed within weighting classes. In other words, respondents are assumed to be a random sample within the cell. In this adjustment, the weighted distribution of respondents is adjusted within a weighting class to equal the distribution of the entire weighting class (i.e., both respondents and nonrespondents).

Weighting class adjustments are effective in reducing nonresponse biases if the weighting classes are internally homogeneous with respect to the response propensity, but as different as possible across classes without unduly inflating sampling variances. Different techniques and procedures can be used to create effective weighting classes. Details of the creation of weighting classes for the *2005 WEOA* are described in the following sections.

For the *2005 WEOA*, weighting classes were used to adjust the base weights for nonresponse in two stages. In the first stage, the base weight was adjusted to account for the circumstance where the eligibility status of some sample members could not be determined. The second stage of adjustment compensated for losses due to eligible members who did not complete the questionnaire. At each stage, the base weights of usable cases (returned and completed questionnaires) were inflated to account for ones that were unusable (not returned, blank, or incomplete questionnaires). The mathematical form of the adjustment is described in the following sections.

Adjusting for Members with Unknown Eligibility

In the first nonresponse adjustment, the base weight is adjusted to account for members with unknown eligibility. As discussed previously, each sampled member was assigned to an appropriate response-status group (i.e., groups *ER, ENR, IN_FR, IN_PR*, or *UNK*). In this stage, members with unknown eligibility (group *UNK*) were assumed to be distributed among the *ER*, *ENR*, and *IN_PR* groups, had it been possible to determine their status. The weights of the members coded as *UNK* were distributed among the members coded as *ER*, *ENR*, and *IN_PR* in the same proportion as observed among the members with known eligibility. In this adjustment, the weights of members coded as *UNK* were not distributed among the frame ineligible members (*IN_FR*). As noted previously, the January 2005 ADMF file was used to identify ineligible members who should have been excluded from the frame (group *IN_FR*). Because all frame ineligible members were identified in the sample, none of their base weights were adjusted in subsequent nonresponse adjustments. In other words, the *IN_FR* cases did not have their weights increased to represent any of the members with unknown eligibility because the entire group could have been identified before data collection.

The first-stage nonresponse adjustment factor was calculated within weighting class c as:

$$
f_c^{A1} = \begin{cases}
\dfrac{\displaystyle\sum_{i \in ER_c} w_i + \sum_{i \in ENR_c} w_i + \sum_{i \in IN\_PR_c} w_i + \sum_{i \in UNK_c} w_i}{\displaystyle\sum_{i \in ER_c} w_i + \sum_{i \in ENR_c} w_i + \sum_{i \in IN\_PR_c} w_i} & \text{If the } i\text{-th sample person classified in weighting class } c \text{ belongs to response group } ER_c, ENR_c, \text{ or } IN\_PR_c. \\[3em]
1 & \text{If the } i\text{-th sample person in class } c \text{ belongs to eligibility group } IN\_FR_c. \\[2em]
0 & \text{If the } i\text{-th sample person in class } c \text{ is in } UNK_c.
\end{cases}
$$

The sums in the numerator of f_c^{A1} extend over the following types of persons in class c: eligible respondents (*ER*), eligible nonrespondents (*ENR*), the proxy-reported ineligible members (*IN_PR*), and members with unknown eligibility (*UNK*). The term w_i is the base weight for the i-th sampled person in class c. The subscript h is omitted for

the sampling stratum since classes crossed strata. However, as described below, the eligibility adjustments and the nonresponse adjustments were almost always made using classes that were the design strata or subdivisions of design strata.

The first nonresponse-adjusted weight w_i^{A1} for a sample member in class c was then computed as:

$$w_i^{A1} = f_c^{A1} w_i \,.$$

Thus, if persons with unknown eligibility accounted for 50% of the weight in class c, the weights of the other units were increased by a factor of 2.

Adjusting for Eligible Nonrespondents

The second nonresponse adjustment increased the adjusted weight of eligible respondents to account for eligible nonrespondents. The second-stage nonresponse adjustment factor for class c was computed as:

$$f_c^{A2} = \begin{cases} \dfrac{\displaystyle\sum_{i \in ER_c} w_i^{A1} + \sum_{i \in ENR_c} w_i^{A1}}{\displaystyle\sum_{i \in ER_c} w_i^{A1}} & \text{If the } i\text{-th sample member in weighting class } c \text{ belongs to response group } ER_c. \\[2em] 0 & \text{If the } i\text{-th sample member sampled in weighting class } c \text{ belongs to response group } ENR_c. \\[2em] 1 & \text{If the } i\text{-th sample member in weighting class } c \text{ belongs to response group } IN\_PR_c \text{ or } IN\_FR_c. \end{cases}$$

The first sum in the numerator of f_c^{A2} for eligible respondents extends over the respondents (group *ER*) in class c; the second extends over the eligible nonrespondents (group *ENR*) in class c; and w_i^{A1} is the previously adjusted weight of the i-th sample member.

The second nonresponse-adjusted weight w_i^{A2}, for the i-th sample member classified in weighting class c was computed as:

$$w_i^{A2} = f_c^{A2} w_i^{A1}.$$

After the two stages of nonresponse adjustment, the nonresponse-adjusted weight for a member in weighting class c becomes

$$w_i^{A2} = f_c^{A2} f_c^{A1} w_i.$$

After the two stages of nonresponse adjustment, members with non-zero weights were those in groups *ER*, *IN_PR*, and *IN_FR*. Eligible nonrespondent members (*ENR*) and members with unknown eligibility (*UNK*) had a zero weight after the adjustments.

Construction of Weighting Classes

The main objective in constructing weighting classes was to group respondents and nonrespondents with similar characteristics into the same adjustment cells. Ideally, the characteristics should be related to both the likelihood of responding to the survey and to survey responses. Each of the characteristics used to create classes must be available for all sampled persons (respondent and nonrespondent). In the *2005 WEOA*, sampling strata were used as the starting point for the creation of weighting classes. The sampling strata were created from variables related to survey response propensity and/or were important reporting domains for survey results (see Table A-1 in Appendix A).

The creation of weighting classes depended primarily on the number of respondents in a sampling stratum. The weighting class corresponded to a sampling stratum when the number of respondents was greater than 30 and smaller than 500. Any stratum with fewer than 30 respondents was combined with another "nearby," or demographically similar, stratum to form a new weighting class. When combining strata, the Service Branch of the member was preserved. This stratification variable was considered a hard boundary that was not crossed when combining strata. The soft boundary variables were station (U.S. versus not U.S.), paygrade group, and race/ethnicity. However, combining strata with different values of race/ethnicity was avoided whenever possible.

There were nine strata with more than 500 respondents. These were subdivided into smaller weighting classes. This subdivision into smaller cells was done using a categorical search algorithm called the Chi-squared Automatic Interaction Detector (CHAID) (Kass, 1980). CHAID attempts to divide the dataset identifying respondents and nonrespondents into groups so that the response rates between cells are as different as possible. Given a set of categorical predictors of response probabilities, CHAID divides the dataset into groups in a stepwise fashion. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictor of response and splits the dataset into categories. Each of those categories is further segmented based on other predictors. Categories of a variable that are not significantly different can be merged together. For the *2005 WEOA*, the merging and splitting continued until no more statistically significant predictors were found or until a user-specified stopping rule was met. No more than six cells were formed within large strata, and each subdivision contained at least 30 respondents.

Dividing the large strata takes advantage of *2005 WEOA* variables not used in stratification. Table 12 lists the variables from the administrative record files that were considered when subdividing large strata, not including the stratification variables.

When the weighting cells contain sufficient cases, and the adjustment factors do not become either too large or too different from each other, the effect on survey variance is often modest. Very large adjustment factors or factors that are much different from others can occur in cells with very high nonresponse rates or with a small number of respondents. Combining cells with few cases to form new cells with at least 30 respondents often compensates for large adjustment factors. However, there are times when cells with more than 30 respondents have a large adjustment factor. If a cell had a large adjustment factor, it was combined with a demographically similar cell to form a new cell with a smaller adjustment factor.

The weighting classes are listed in Table B-1 in Appendix B. These classes were used for both the first and second stages of nonresponse adjustment. The table also lists the adjustment factors, f_c^{A1} and f_c^{A2}, for each class.

Table 12.
Member Characteristics Considered for Creation of Nonresponse Weighting Classes Within Sampling Strata With 500 or More Respondents

Description	Level	Values
Paygrade	1	E1
	2	E2
	. . .	. . .
	9	E9
	10	Unknown Enlisted
	11	W1
	12	W2
	. . .	. . .
	15	W5
	16	Unknown Warrant Officer
	17	O1
	18	O2
	. . .	. . .
	22	O6
	23	Unknown Officer
	24	Unknown paygrade
Age Groups	1	17, 18 years old
	2	19, 20 years old
	3	21, 22 years old
	4	23, 24 years old
	. . .	. . .
	22	59, 60 years old
	23	61, 62 years old
	24	63, 64 years old
	25	65, 66 years old
	26	Unknown

Table 12.   (Continued)

Description	Level	Values
Region	1	US & US Territories
	2	Europe
	3	Other
	4	Asia & Pacific Islands
	5	Unknown
Length Away Due to Occupation	1	0.321 – 1.06 months
	2	1.07 – 1.82 months
	3	1.83 – 2.58 months
	4	2.59 – 3.34 months
	5	3.35 – 4.10 months
	6	4.11 – 4.86 months
	7	Unknown
Dual Spouse Status	1	No Dual Service spouse
	2	Dual Reserve/Guard spouse
	3	Dual Active spouse
	4	Unknown, NA
Family Status	1	Single w/children
	2	Single w/o children
	3	Married w/ children
	4	Married w/o children
	5	Unknown child count
Level of Education	1	No College
	2	Some College
	3	Four-year degree
	4	Graduate/Professional degree
	5	Unknown
Marital Status	1	Married
	2	Not Married
	3	Unknown
Detailed Race	1	White
	2	Black
	3	American Indian
	4	Asian
	5	Hawaiian/Pacific Islander
	6	Multi-race
	7	Unknown
Gender	1	Male
	2	Female

Note. Stratification variables are not presented in the table.

Raking Adjustment

As indicated above, the final step in weighting, raking, is intended to increase the precision of survey population estimates by benchmarking them to known population values. The nonresponse-adjusted weights were raked to force weighted sample estimates to equal known population totals (Brackstone & Rao, 1976; Wolter, 1985; and Kalton & Flores Cervantes, 2003). The mechanics of the raking weight adjustment is summarized below.

The population was partitioned, based on the first raking dimension, into groups denoted by $U_1, \ldots, U_G$. The groups are, by definition, mutually exclusive and cover the entire population. Let N_g be the size of U_g, so that $N = \sum_{g=1}^{G} N_g$. The eligible respondents (*ER*) and ineligible members[6] (*IN_PR*) in the sample were also partitioned into groups $s_1, \ldots, s_G$. The expression for the initial weighting adjustment factor for all the units classified in cell g is

$$\widetilde{f}_g^R = \frac{N_g}{\sum_{i \in s_g} w_i^{A2}}.$$

The raked weight $\widetilde{w}_i^R$, for the i-th sample member classified in cell g of the first raking dimension was then computed as:

$$\widetilde{w}_i^R = \widetilde{f}_g^R w_i^{A2}, i \in s_g.$$

A similar adjustment was then made after classifying the sample based on the second raking dimension. Successively adjusting the weights through the third up to the last dimension (K) constitutes the first iteration of the process. The adjusted weights for $i=2$ through K result in and estimate of the sum of weights for members classified by dimension 1. The adjustments for dimensions 1 to K are carried out again beginning with the adjusted weights from the first iteration. The iterative process continues until the sum of the weights for each raking dimension is acceptably close to the corresponding control total. The final raked weight w_i^R, for the i-th sample person was then computed as:

$$\widetilde{w}_i^R = \widetilde{f}_g^R w_i^{A2}, i \in s_g$$

where $\widetilde{f}_g^R$ is the product of the iterative adjustments applied to the i-th sample member.

Control Totals and Raking Dimensions

The population or control totals for the raking dimensions were computed using the updated frame created from the January 2005 ADMF. The control totals were

---

[6] Ineligible members coded as *IN_FR* were not raked because they were excluded from control totals.

computed excluding all ineligible members. The variable F_ELIG was created for all records in the frame as shown in Table 2. The total population used in raking included 1,319,408 members.

Some sample members who were eligible in the *2005 WEOA* sampling frame were reported by themselves or by proxies as ineligible after the creation of the frame. Those members received a separate ineligibility code (*IN_PR*) as noted earlier. Existence of such persons in the sample was evidence that the sampling frame contained members who became ineligible after the frame was created. Consequently, sample persons coded as eligible respondents (*ER*) and ineligibles (*IN_PR*) were both included in raking.

For the *2005 WEOA*, in addition to the stratification variables, the following variables were considered for creating raking dimensions and cells in addition to the stratification variables:

- Gender (Male, Female)

- Detailed Paygrade (Enlisted E1, E2, . . . , E9; Warrant Officer W1, W2, . . . , W5; Commissioned Officer O1, O2, . . . , O6);

- Detailed Region (U.S. North, U.S. South, U.S. West, Europe, Asia/Pacific Islands, Other, Unknown);

- Age categories (17-24 years old, 25-29 years old, 30-34 years old, 35-39 years old, 40-44 years old, 45-49 years old, 50 years old or older);

- Educational Attainment (Unknown, no college, some college, four-year degree, grad/professional degree); and

- Marital Status (Married, Not Married, Unknown).

Table 13 shows the final raking dimensions used in the *2005 WEOA*.

Table 13.
Definition of the Dimensions (With Number of Categories) Used in Raking

Dimension	Definition
1	Service Branch (5) by gender (2) by age (7)
2	Service Branch (5) by paygrade group (5)
3	Service Branch (5) by race/ethnicity (8)
4	Service Branch (5) by detailed region (7)
5	Detailed paygrade group (20)

Cells for dimensions with fewer than 100 respondents were collapsed to form new cells. Initially, the weights were raked using Dimensions 1 to 4. These dimensions reflected the variables used for stratification, with additional levels previously not used in the creation of the sampling strata. For example, the stratification levels of race/ethnicity included Asian and Hawaiian/Pacific Islander as a single race category, while in raking these races were separated into two categories. Variables for age group and gender that were not used in stratification were used to create Dimension 1. These additional variables and the more detailed levels of the stratification variables were useful in reducing residual bias not adjusted by the nonresponse adjustment, because these detailed levels were not used in the creation of weighting classes.

After the weights were raked, ratios of totals from the frame to the sum of weights (base, nonresponse-adjusted, and raked weights) were computed and examined to determine if there was any distortion or bias introduced to the weights after each adjustment. The ratios showed that younger members and members in the enlisted paygrades were underestimated. An additional dimension (Dimension 5) was included and the ratios were recomputed and evaluated. Dimension 5 was included because there were differential response rates within individual paygrades that were not accounted for by the other dimensions or by the nonresponse adjustments.

The categories and control totals for each of these variables are listed in Tables C-1 to C-5 in Appendix C. Note that by creating dimensions for raking that are crosses of two or more individual variables, some degree of interaction among the variables is accounted for when creating the raked weight.

Table 14 shows the overall raking ratios for selected variables before the beginning of raking. The overall raking factor was computed as the ratio of the total from the frame to the sum of weights before raking. The ratio was a measure of how effective the nonresponse adjustments were at removing the bias for these variables. For variables used in stratification, the overall raking ratio was close to 1.0. This was expected because the weighting cells were created using the sampling strata.

Ratios different than 1.0 were indicative of residual biases not removed by the nonresponse adjustments. Ratios greater than 1.0 indicate that a group is underrepresented because the sum of the weights prior to raking for the group was less than the control total. On the other hand, a ratio less than 1.0 means that the weights for a group were likely over adjusted during the nonresponse-adjustment because the sum of weights prior to raking was greater than the control total. These biases were somewhat larger for females, some age groups (i.e., 17 to 24 years, 30 years or older), some paygrades (enlisted), American Indians, members not stationed in the U.S. or Europe, married members, and members with some college. These biases reflect that the weighting classes created using stratification variables were not optimal to reduce the nonresponse bias for these characteristics. The variables in Table 13 were used to create the raking dimensions; therefore, these biases were removed after raking because the sums of weights were forced to match control totals for the dimensions.

Table 14.
Overall Raking Ratios for Selected Variables Prior to Raking

Variable	Overall Raking Ratio
Service Branch	
Army	1.00
Navy	1.00
Marine Corps	1.00
Air Force	1.00
Coast Guard	1.01
Gender	
Male	1.03
Female	0.84
Age Group	
17 to 24 years, unknown	1.12
25 to 29 years	0.97
30 to 34 years	0.90
35 to 39 years	0.93
40 to 44 years	0.94
45 to 49 years	0.92
50 years or older	0.90
Detailed Paygrade	
E1-E2	1.65
E3	1.12
E4, Unknown Enlisted	0.92
E5	1.01
E6	0.90
E7	0.97
E8	0.97
E9	0.81
W1-W2	1.00
W3	0.96
W4-W5	0.90
O1	1.02
O2	1.06
O3	1.01
O4, Unknown Officer	0.97
O5	0.98
O6	0.95
Race/ethnicity	
White	1.00
Black	1.00
Hispanic	1.00
Asian	0.99
American Indian	1.22
Hawaiian/Pacific Islander	1.01
Multiple Race	0.92
Unknown	0.99

Table 14.   (Continued)

Variable	Overall Raking Ratio
Location	
U.S. - North	0.93
U.S. - South	1.00
U.S. - West	1.01
Europe	0.98
Asian/Pacific Islands	0.89
Other	1.77
Unknown	0.78
Marital Status	
Married, Unknown	1.10
Not married	0.94
Educational Attainment	
No college, Unknown	1.03
Some college	0.88
Four-year degree	0.95
Grad/Professional degree	0.97

Final Weights

Table 15 shows the sample counts and sum of weights after the weighting adjustments for the entire sample and each disposition code. After raking, the cases with non-zero weights were those coded as *ER* and *IN_PR*. These cases have final weights equal to

$$\widetilde{w}_i^R = F_c^{A1} F_c^{A2} \widetilde{F}_g^R w_i$$

where $F_c^{A1}, F_c^{A2}, \widetilde{F}_g^R$, and w_i are defined in the Weighting Procedures section of the report. Cases coded as *ENR, IN_FR*, and *UNK* have zero weights.

Table 15.
Sample Counts and Sum of Weights After Weighting Adjustments

Weighting Step	Total
1 Base weight	
1.1 Sample size	91,024
1.2 Sum of weights	1,376,874
1.3 Coefficient of Variation[a]	79.84

Table 15.   (Continued)

Weighting Step		Total
1.4	Number of records by disposition code	91,024
	a. Eligible respondents (*ER*)	32,299[b]
	b. Eligible nonrespondents (*ENR*)	3,622
	c. Ineligible members—by proxy (*IN_PR*)	189
	d. Ineligible members—on frame (*IN_FR*)	3,609
	e. Member with unknown eligibility—(*UNK*)	51,305
2	Adjustment for unknown eligibility	
2.1	Sum of weights before adjustment	1,376,874
	a. Eligible respondents (*ER*)	508,817
	b. Eligible nonrespondents (*ENR*)	52,725
	c. Ineligible members—by proxy (*IN_PR*)	2,700
	d. Ineligible members—on frame (*IN_FR*)	56,880
	e. Member with unknown eligibility—(*UNK*)	755,751
2.2	Sum of weights after adjustment	1,376,874
	a. Eligible respondents (*ER*)	1,161,289
	b. Eligible nonrespondents (*ENR*)	152,098
	c. Ineligible members—by proxy (*IN_PR*)	6,606
	d. Ineligible members—on frame (*IN_FR*)	56,880
	e. Member with unknown eligibility—(*UNK*)	0
2.3	Coefficient of Variation - positive weights only	97.49
3	Nonresponse adjustment	
3.1	Sum of base weights before adjustments	1,376,874
	a. Eligible respondents (*ER*)	1,161,289
	b. Eligible nonrespondents (*ENR*)	152,098
	c. Ineligible members—by proxy (*IN_PR*)	6,606
	d. Ineligible members—on frame (*IN_FR*)	56,880
	e. Member with unknown eligibility—(*UNK*)	0
3.2	Sum of base weights after adjustments	1,376,874
	a. Eligible respondents (*ER*)	1,313,387
	b. Eligible nonrespondents (*ENR*)	0
	c. Ineligible members—by proxy (*IN_PR*)	6,606
	d. Ineligible members—on frame (*IN_FR*)	56,880
	e. Member with unknown eligibility—(*UNK*)	0
3.3	Coefficient of Variation - positive weights only	102.50
4	Raking adjustment	
4.1	Sum of weights before adjustment of records include in raking	1,319,994
	a. Eligible respondents (*ER*)	1,313,387
	b. Ineligible members—by proxy (*IN_PR*)	6,606
4.2	Sum of weights not included in raking	56,880
	a. Ineligible members—on frame (*IN_FR*)	56,880
	b. Member with unknown eligibility—(*UNK*)	0
4.3	Control total	1,319,408
4.4	Overall Raking factor	0.99

Table 15.   (Continued)

Weighting Step		Total
4.5	Sum of weights after adjustment	1,319,408
	a. Eligible respondents (*ER*)	1,312,934
	b. Eligible nonrespondents (*ENR*)	0
	c. Ineligible members—by proxy (*IN_PR*)	6,474
	d. Ineligible members—on frame (*IN_FR*)	0
	e. Member with unknown eligibility—(*UNK*)	0
4.6	Coefficient of Variation - positive weights only	110.20

[a] The coefficients of variation (CVs) are presented to assess the increase of variability of the weights after each weighting adjustment. Most DMDC sample designs, including the *2005 WEOA*, use differential sampling rates and deep stratification. This type of design leads to high CV values for the full sample and is more appropriate for producing estimates for small domains.
[b] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

The 1996 Equal Opportunity Survey

The previous survey that addressed topics related to the frequency and effects of racial/ethnic harassment and discrimination experienced by active duty military was administered by DMDC in 1996. This survey is referred to as the *1996 Armed Forces Equal Opportunity Survey* (*1996 EOS*). The *1996 EOS* was a mail survey with a pre-notification letter, two waves of questionnaire mailings, and a reminder/thank you letter following the first wave of questionnaire mailing.

The *1996 EOS* population of interest included members of the National Guard and Reserve in active duty assignments [i.e., Active Guard Reserve (AGR) and Navy Training and Administration of Reserve (TAR)] for at least 179 days, in addition to Army, Navy, Marine Corps, Air Force, and Coast Guard members below the rank of admiral or general with at least six months of active duty service. The *2005 WEOA* included members in active duty and no data was collected for National Guard and Reserve members in active duty assignments. As a result, direct comparisons between estimates from the *2005 WEOA* and *1996 EOS* are not possible due to differences in the survey populations. However, comparisons of domains common to both surveys are possible. Common domains can be created using the variable CSERVICE (Constructed Service) from the *1996 EOS* confidential dataset. The values of CSERVICE are shown below in Table 16.

In order to produce comparisons between the *1996 EOS* and the *2005 WEOA*, the *1996 EOS* data files should be subset to include records where CSERVICE ≠ 6.

Table 16.
Values of the variable CSERVICE (Constructed Service) from the 1996 EOS

Values	Description	Number of Cases	Percentage
1	Army	24,595	32.04
2	Navy	15,892	20.71
3	Marine Corps	12,363	16.11
4	Air Force	14,230	18.54
5	Coast Guard	6,124	7.98
6	AGR/TAR	3,550	4.63
Total		76,754	100.00

Differences in the Development of Weights of the 1996 EOS and the 2005 WEOA

Analytical weights were created for the *1996 EOS* and they reflected the probability of selection of the member, a nonresponse adjustment factor, and a poststratification factor. Logistic models of unit nonresponse propensity were used to generate nonresponse adjustment factors. The weights were poststratified to the population as of the beginning of the data collection period.

There were three main differences between the development of the weights for the *1996 EOS* and the development of the *2005 WEOA* weights. The differences included the assignment of final disposition codes, the method of creating nonresponse adjustment cells, and finally, the method used to benchmark the analysis weights to population frame control totals. The following paragraphs describe these differences.

The method for assigning final disposition codes in the *1996 EOS* was not the same as that used for the *2005 WEOA*. The difference was in the treatment of eligible nonrespondents (*ENR*) and ineligible members identified though administrative files (*IN_FR*). For the *1996 EOS*, *ENR* members identified as eligible but who returned incomplete and unusable surveys were treated as members with unknown eligibility (*UNK*). The *IN_FR* members were treated as self- or proxy-reported members (*IN_PR*). This assumption was reflected in the two steps used to create the analytical weights in the *1996 EOS*. In the first step, the base weights were adjusted to account for members with unknown eligibility (*UNK*). In the second step, the nonresponse-adjusted weights were poststratified to control totals.

A second difference in the development of the weights for the two surveys was the method for creating nonresponse adjustment cells. In the *1996 EOS* the weights were adjusted using factors from logistic models that estimated unit response propensity. The *2005 WEOA* method used weighting classes where most of the classes corresponded to sample strata. In other words, the nonresponse adjustments in the *2005 WEOA* relied mainly on the variables used for stratification and additional variables were used only in a few large strata.

The last difference was in the type of adjustment used to benchmark the nonresponse-adjusted weights to the frame control totals. In the *1996 EOS* the weights were poststratified to control totals defined by Service (Army, Navy, Marine Corps, Air Force, Coast Guard, and AGR/TARS), and by detailed race/ethnicity (Hispanic, White Non-Hispanic, African American Non-Hispanic, American Indian, Asian/Pacific Islander, and Other). The *2005 WEOA* weights were adjusted using raking. The raking dimensions were described in previous sections. In the *2005 WEOA*, the raking adjustment not only benchmarked the sum of weights to known control totals but also reduced any residual biases not accounted for in the nonresponse adjustments.

Because of the small number of sample members who were handled differently, the impact of different weighting methodologies on the estimates is minimal. However, because the frame ineligibles (*IN_FR*) are removed from the sample and control totals in the *2005 WEOA*, the estimates of ineligible members in the sample are smaller.

Variance Estimation

Variance estimation procedures are developed to account for the sample design and estimators employed in a complex survey. Using these procedures, analysts can appropriately reflect factors, such as sample selection in multiple stages, and the use of differential sampling rates to oversample a targeted subpopulation in estimates of sampling error. The two main methods for estimating variances from a complex survey are known as linearization (or Taylor series variance estimation) and replication. Wolter (1985) and Shao (1996) describe the theory and applications of these methods. The special variables needed to compute variances using these methods were created for the *2005 WEOA*. Depending on the analysis, data users can choose either method to compute the estimates of variance. A general description of these methods is included in Appendix D.

For complex sample surveys, such as the *2005 WEOA*, the computation of sampling errors requires specialized software. Many standard statistical software packages assume a simple random sample when computing estimates of variance. However, estimates of variance from these packages can seriously understate the true variability of the survey estimates. In recent years, specialized commercial software has been developed to analyze data from complex surveys (Lepkowski & Bowles, 1996; Cohen, 1996; Broene & Rust, 1998). Appendix F also includes a description of statistical software for variance estimation for the *2005 WEOA*.

Location, Completion, and Response Rates

Response rates are generally used to measure the quality of a survey. Although the use of response rates as a single measure of the quality of a survey is overstated, they do provide valuable information on the success of the survey in representing the population sampled (Madow, Nisselson, & Olkin, 1983).

CASRO has pointed out that varying operational definitions of response rates can lead to misleading conclusions. In an effort to standardize the operational definition and

computation of response rates in surveys, CASRO published guidelines and recommendations in 1982 (Council of American Survey Research Organizations, 1982). Beginning in 1995, DMDC standardized its methods for calculating response rates using procedures patterned after those advocated by CASRO. More specifically, the DMDC procedures closely follow CASRO's Sample Type II design.

The main objective of this section is to present response rates that can be used by analysts of the *2005 WEOA* data to better understand how well the member population is represented. To accomplish this goal, response rates are weighted so that they are an estimate of the proportion of the population responding (i.e., response propensity in the population). For example, because the sample was selected with differing sampling rates by sampling strata, the response rates are weighted so each stratum accounts for its appropriate fraction when the total response rate is reported. Observed or unweighted response rates are useful for monitoring the survey during data collection. However, when different subpopulations are either undersampled or oversampled, weighted response rates are needed to compare response rates for different sample groups.

Table 17 shows the weighted and unweighted location, completion and response rates computed for the *2005 WEOA*. The location rate (*LR*) is defined as the proportion of eligible sample members who were locatable. The completion rate (*CR*) is defined as the proportion of the located sample who returned usable surveys, while the response rate (*RR*) is defined as the proportion of eligible sample members who returned usable surveys. The response rate (*RR*) is computed as the product of the location rate (*LR*) and the completion rate (*CR*); that is:

$$RR = LR * CR.$$

These rates are adjusted for ineligible members to account for the unknown eligibility of some members, as described in previous sections.

Table 17.
Location Rates, Response Rates, and Completion Rates

Type of Rate	Observed Rate	Weighted Rates
Location (LR)	87.75%	88.47%
Completion (CR)	42.33%	43.78%
Response (RR)	37.14%	38.73%

The location, completion and response rates can be also expressed as ratios of the adjusted located sample (N_L), the adjusted eligible sample (N_E), and the usable responses (N_R) as follows:

The *location rate* is defined as

$$LR = \frac{\text{Adjusted located sample}}{\text{Adjusted eligible sample}} = \frac{N_L}{N_E}.$$

The *completion rate* is defined as

$$CR = \frac{\text{Usable responses}}{\text{Adjusted located sample}} = \frac{N_R}{N_L}.$$

The *response rate* is defined as

$$RR = \frac{\text{Usable responses}}{\text{Adjusted eligible sample}} = \frac{N_R}{N_E}.$$

The rates in Table 17 were computed using the information from Table 18 that shows the weighted and unweighted distribution of the located, eligible, and usable samples for the *2005 WEOA*. In this table, the adjusted eligible sample and adjusted locatable sample were computed by subtracting the estimated number of ineligible members from the count of members who were not located or who did not return the survey.

Table 18.
Frequency Counts and Percents of the Final Sample Relative to the Drawn Sample

Description	Sampled Cases n	Sampled Cases %	Sums of Base Weights	Sums of Base Weights %
Drawn Sample & Population	**91,024**	**100.00%**	**1,376,874**	**100.00%**
Total Ineligible	**3,798**	**4.17%**	**59,580**	**4.33%**
Ineligible on Master File	3,609	3.96%	56,880	4.13%
Self-reported ineligible	189	0.21%	2,700	0.20%
Total Eligible Sample	**87,226**	**95.83%**	**1,317,294**	**95.67%**
Total Not Located Sample	**10,710**	**11.77%**	**152,209**	**11.05%**
Not located - estimated ineligible	56	0.06%	728	0.05%
Not located - estimated eligible	10,654	11.70%	151,481	11.00%
Total Located Sample	**76,516**	**84.06%**	**1,165,085**	**84.62%**
Total Nonrespondents	**44,217**	**48.58%**	**656,268**	**47.66%**

Table 18.   (Continued)

Description	Sampled Cases n	Sampled Cases %	Sums of Base Weights	Sums of Base Weights %
Returned blank	24	0.03%	390	0.03%
Skipped key questions	3,101	3.41%	45,570	3.31%
Requested removal from survey mailings	497	0.55%	6765	0.49%
Did not returned a survey (DNR)	40,595	44.60%	603,543	43.83%
DNR - estimated ineligible	212	0.23%	2,888	0.21%
DNR - estimated eligible	40,383	44.36%	600,655	43.62%
Usable Responses	**32,299**[a]	**35.48%**	**508,817**	**36.95%**

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

Details of the computation of the rates are described in Appendix E. Weighted and unweighted rates are also reported for the full sample and categories of Service Branch, location, paygrade group, race/ethnicity, gender, education and marital status shown in Table 19. In this table, base weights were used in computing the weighted rates. Table E-3 in Appendix E lists the same rates by sampling strata.

In recent years, use of the American Association for Public Opinion Research (AAPOR) guidelines and definitions for computing rates has grown in popularity (AAPOR, 2004). The CASRO rate definitions used in the 2005 WEOA have corresponding AAPOR definitions. The response rate (*RR*) as defined above corresponds to AAPOR's response rate 3 (*RR3*) that uses the estimate of proportion of cases of unknown eligibility who are actually eligible. The estimate of eligible cases among the cases with unknown eligibility is based on the observed proportion of eligible cases in the sample as is described in Appendix E. The location rate (*LR*) is equivalent to AAPOR's contact rate 2 (*CON2*) and includes in the denominator only the estimated eligible cases among the undetermined cases. Finally, the completion rate (*CR*) corresponds to AAPOR's cooperation rate 1 (*COOP1*) also known as the minimum cooperation rate. These equivalencies allow the equation of CASRO and AAPOR response rates. In the present case

$$RR = LR * CR = CON2 * COOP1 = RR3.$$

Table 19.

Unweighted and Weighted Location, Completion, and Response Rates for the Full Sample and Categories of Service Branch, Location, Paygrade Group, Race/Ethnicity, Gender, Education and Marital Status

Group	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Unweighted Rate			Weighted Rate		
				Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
Full Sample	86,957	76,304	32,299[a]	87.70	42.30	37.10	88.50	43.80	38.70
Service Branch									
Army	31,536	27,717	11,830	87.90	42.70	37.50	85.70	38.00	32.50
Navy	19,094	16,709	7,628	87.50	45.70	40.00	88.00	46.00	40.50
Marine Corps	18,099	14,744	4,101	81.50	27.80	22.70	82.00	28.20	23.10
Air Force	15,171	14,229	7,394	93.80	52.00	48.70	94.60	53.70	50.80
Coast Guard	3,051	2,899	1,346	95.00	46.40	44.10	95.80	51.60	49.40
Location									
U.S. (Including Territories)	64,640	56,781	24,574	87.80	43.30	38.00	88.50	44.40	39.30
Europe	9,092	8,379	3,592	92.20	42.90	39.50	92.20	44.40	40.90
Asia/Pacific Islands	9,728	8,369	3,344	86.00	40.00	34.40	86.90	40.00	34.80
Other	3,365	2,658	735	79.00	27.70	21.80	78.80	29.70	23.40
Unknown	138	121	54	87.70	44.60	39.10	87.70	44.60	39.10
Paygrade Group									
Enlisted E1 to E4	40,713	32,766	8,251	80.50	25.20	20.30	80.80	26.90	21.70
Enlisted E5 to E9	21,852	20,281	9,955	92.80	49.10	45.60	93.50	51.20	47.90
Warrant Officer W1 to W5	2,661	2,562	1,496	96.30	58.40	56.20	96.00	59.40	57.00
Officer O1 to O3	13,495	12,596	6,899	93.30	54.80	51.10	94.20	58.10	54.70
Officer O4 to O6	8,236	8,099	5,698	98.30	70.40	69.20	98.40	71.10	70.00

Table 19. **(Continued)**

Group	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Unweighted Rate			Weighted Rate		
				Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
Race/Ethnicity									
White	40,770	36,456	17,412	89.40	47.80	42.70	89.10	46.10	41.10
Black	15,990	13,891	5,120	86.90	36.90	32.00	87.50	37.60	32.90
Hispanic	15,805	13,427	4,544	85.00	33.80	28.80	86.50	38.70	33.50
Native American	4,798	4,057	1,425	84.60	35.10	29.70	84.30	36.00	30.40
Asian/Pacific Islander	6,711	5,874	2,567	87.50	43.70	38.30	88.00	45.70	40.20
Other	335	296	139	88.40	47.00	41.50	88.90	45.60	40.50
Unknown	2,553	2,306	1,092	90.30	47.40	42.80	89.60	45.80	41.00
Gender									
Male	73,956	64,617	27,081	87.40	41.90	36.60	88.10	43.30	38.20
Female	13,005	11,689	5,218	89.90	44.60	40.10	90.60	46.30	41.90
Education									
No College	54,955	46,097	14,975	83.90	32.50	27.20	86.00	37.00	31.90
Some College	5,311	4,942	2,691	93.10	54.50	50.70	94.30	58.50	55.20
Four-year degree	14,959	14,006	7,890	93.60	56.30	52.70	94.00	59.00	55.50
Grad/Prof degree	7,121	6,975	4,774	97.90	68.40	67.00	98.00	69.30	67.90
Unknown	4,627	4,295	1,969	92.80	45.80	42.60	93.50	49.10	45.90
Marital Status									
Married	43,737	40,964	20,649	93.70	50.40	47.20	94.00	50.80	47.80
Not married	43,199	35,326	11,650	81.80	33.00	27.00	82.20	34.70	28.50
Unknown									

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

References

American Association for Public Opinion Research. (2004). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* (3rd ed.). *Lenexa, KS: AAPOR.*

Armed forces equal opportunity survey [Data file and documentation on CD-ROM]. Arlington, VA: DMDC.

Brick, J. M. & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5: 215-238.

Brackstone, G. J., & Rao, J. N. K. (1976). Raking ratio estimators. *Survey Methodology*, 2: 63-69.

Broene, P., & Rust, K. (1998). Strengths and limitations of using SUDAAN, Stata, and WesVarPC for computing variances from NCES data sets. Working Paper, U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Cohen, S. (1996). *An evaluation of alternative PC-based software packages developed for the analysis of complex survey data.* Rockville, MD: Agency for Health Care Policy and Research.

Council of American Survey Research Organizations (1982). *On the definition of response rates* (special report of the CASRO task force on completion rates, Lester R. Frankel, Chair). Port Jefferson, NY: Author.

DMDC (2004). *2003 Survey of retired military: Administration, datasets, and codebook* (Report No. 2004-009). Arlington, VA: DMDC.

Elig, T. W., Riemer, R. A., Simmons, R. O., & Valliant, R. (2002). *2000 Reserve component surveys: Statistical methodology report* (Report No. 2002-003). Arlington, VA: DMDC.

George, B. J. & Kroeger, K. (Ed.). (2002). *Weighting report for the armed forces 2002 workplace and gender relations survey.* Arlington, VA: DMDC.

Kalton, G., & Flores Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics, 19,* 81 – 97.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119-127.

Kavee, J. D., & Mason, R. E. (1997). *DMDC sample planning tool* (Report No. 97-028). Arlington VA: DMDC.

Kish, L. (1965). *Survey sampling.* New York. John Wiley & Sons, Inc.

Lepkowski, J., & Bowles, J. (1996). Sampling error software for personal computers. *The Survey Statistician*, 35, 10-16.

Madow, W., Nisselson, H., & Olkin, I. (1983). *Incomplete data on sample surveys, volume 1 – Report and case studies*. New York: Academic Press.

Mason, R. E., Kavee, J. A., Wheeless, S. C., George, B. J., Riemer, R. A., & Elig, T. W (1996). *The 1995 armed forces sexual harassment survey: Statistical methodology report* (Report No. 96-016). Arlington VA: DMDC.

Research Triangle Institute (2001). *SUDAAN® user's manual, (Release 8.0)*. Research Triangle Park, NC: Author.

Riemer, R. A. (Ed.) (2004). *2004 Workplace and gender relations survey of reserve component members: Statistical methodology report* (Report No. 2004-019). Arlington, VA: DMDC.

Riemer, R. A. & Lamoreau, D. P. (2002). *1996 Retired military personnel survey: Statistical methodology report* (Report No. 1998-014). Arlington, VA: DMDC.

Rust, K. F. (1986). Efficient replicated variance estimation. *1986 Proceedings of the Section on Survey Research Methods* (pp. 81–87). Alexandria, VA: American Statistical Association.

Rust, K. F., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research, 5*, 282–310.

SAS Institute, Inc. (2001). *SAS/STAT® user's guide, version 8.0.* Cary, NC: SAS Institute, Inc., 1999, 3884pp.

Shao, J. (1996). Resampling methods in sample surveys, (with discussion). *Statistics*, 27, 203-254.

Skinner, C., Holt, D., & Smith, T. (Eds.) (1989). *Analysis of complex surveys*. New York: Wiley.

Valliant, R. (2002). The effect of multiple weighting steps on variance estimation. *2002 Proceedings of the Section on Survey Research Methods* (pp. 3547-3552). Alexandria, VA: American Statistical Association.

Westat. (2000). *WesVar™ 4.0 user's guide*. Rockville, MD: Westat.

Wheeless, S. C., Mason, R. E., & Kavee, J. D. (1997). *Armed forces 1996 equal opportunity survey: Statistical methodology report* (Report No. 97-025). Arlington, VA: DMDC.

Wolter, K. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.

Wright, L. C., Elig, T. W., Flores-Cervantes, I., George, B. J., & Valliant, R. (2000). *1999 Survey of active duty personnel: Statistical methodology report* (Report No. 2000-010). Arlington, VA: DMDC.

Appendix A.
Sample Selection Tables

Table A-1.
Stratum Definition for the 2005 WEOA

Stratum Number	Service	Region	Pay Group	Race/ Ethnicity	Stratum Size	Sample Allocation	Sample Size	Percent Sampled
1	Army	United States	E1-E3	White	51,035	474	2,407	4.7
2	Army	United States	E1-E3	Black	12,916	94	994	7.7
3	Army	United States	E1-E3	Hispanic	9,754	93	391	4.0
4	Army	United States	E1-E3	Native American	806	57	297	36.8
5	Army	United States	E1-E3	Asian/Pacific Islander	2,690	45	161	6.0
6	Army	United States	E1-E3	Other	587	9	37	6.3
7	Army	United States	E4	White	52,623	530	2,132	4.1
8	Army	United States	E4	Black	19,239	183	1,171	6.1
9	Army	United States	E4	Hispanic	10,859	110	385	3.5
10	Army	United States	E4	Native American	843	67	276	32.7
11	Army	United States	E4	Asian/Pacific Islander	2,821	52	155	5.5
12	Army	United States	E4	Other	1,598	25	82	5.1
13	Army	United States	E5-E6	White	51,860	600	1,341	2.6
14	Army	United States	E5-E6	Black	29,599	393	1,095	3.7
15	Army	United States	E5-E6	Hispanic	10,161	116	244	2.4
16	Army	United States	E5-E6	Native American	718	77	175	24.4
17	Army	United States	E5-E6	Asian/Pacific Islander	2,149	49	90	4.2
18	Army	United States	E5-E6	Other	3,847	74	143	3.7
19	Army	United States	E7-E9	White	19,293	231	353	1.8
20	Army	United States	E7-E9	Black	14,307	219	363	2.5
21	Army	United States	E7-E9	Hispanic	3,119	33	50	1.6
22	Army	United States	E7-E9	Native American	318	41	65	20.4
23	Army	United States	E7-E9	Asian/Pacific Islander	560	14	19	3.4
24	Army	United States	E7-E9	Other	1,949	44	63	3.2
25	Army	United States	W1-O6	White	45,889	4,347	6,938	15.1
26	Army	United States	W1-O6	Black	7,936	942	1,802	22.7
27	Army	United States	W1-O6	Hispanic	3,105	372	604	19.5
28	Army	United States Not United States	W1-O6	Native American	379	178	307	81.0

Table A-1.   (Continued)

Stratum Number	Service	Region	Pay Group	Race/ Ethnicity	Stratum Size	Sample Allocation	Sample Size	Percent Sampled
29	Army	United States	W1-O6	Asian/Pacific Islander	1,695	203	299	17.6
30	Army	United States	W1	Other	1,892	282	432	22.8
31	Army	Not United States	E1-E3	White	15,025	146	741	4.9
32	Army	Not United States	E1-E3	Black	4,067	30	317	7.8
33	Army	Not United States	E1-E3	Hispanic	3,112	283	1,188	38.2
34	Army	Not United States	E1-E3	Native American	254	18	94	37.0
35	Army	Not United States	E1-E3	Asian/Pacific Islander	968	148	530	54.8
36	Army	Not United States	E1-E3	Other	365	6	24	6.6
37	Army	Not United States	E4	White	13,781	151	587	4.3
38	Army	Not United States	E4	Black	5,275	56	266	5.0
39	Army	Not United States	E4	Hispanic	3,012	222	776	25.8
40	Army	Not United States	E4	Native American	214	17	70	32.7
41	Army	Not United States	E4	Asian/Pacific Islander	951	178	531	55.8
42	Army	Not United States	E4	Other	462	7	23	5.0
43	Army	Not United States	E5-E6	White	13,501	162	362	2.7
44	Army	Not United States	E5-E6	Black	8,957	122	320	3.6
45	Army	Not United States	E5-E6	Hispanic	2,826	298	625	22.1
46	Army	Not United States	E5-E6	Native American	205	22	50	24.4
47	Army	Not United States	E5-E6	Asian/Pacific Islander	816	187	345	42.3
48	Army	Not United States	E5-E6	Other	1,103	21	40	3.6
49	Army	Not United States	E7-E9	White	3,812	48	76	2.0
50	Army	Not United States	E7-E9	Black	3,785	55	110	2.9
51	Army	Not United States	E7-E9	Hispanic	721	93	140	19.4
52	Army	Not United States	E7-E9	Native American	91	12	19	20.9

53	Army	Not United States	E7-E9	Asian/Pacific Islander	177	46	63	35.6
54	Army	Not United States	E7-E9	Other	511	11	16	3.1
55	Army	Not United States	W1-O6	White	10,462	1,077	1,836	17.5
56	Army	Not United States	W1-O6	Black	1,855	236	447	24.1
57	Army	Not United States	W1-O6	Hispanic	694	121	196	28.2
58	Army	Not United States	W1-O6	Asian/Pacific Islander	568	153	225	39.6
59	Army	Not United States	W1-O6	Other	455	63	96	21.1
60	Navy	United States	E1-E3	White	43,021	436	2,417	5.6
61	Navy	United States	E1-E3	Black	16,397	119	946	5.8
62	Navy	United States	E1-E3	Hispanic	5,006	45	257	5.1
63	Navy	United States	E1-E3	Native American	3,398	127	580	17.1
64	Navy	United States	E1-E3	Asian/Pacific Islander	3,472	45	166	4.8
65	Navy	United States	E1-E3	Other	2,285	38	101	4.4
66	Navy	United States	E4	White	29,314	333	1,456	5.0
67	Navy	United States	E4	Black	10,925	103	624	5.7
68	Navy	United States	E4	Hispanic	6,727	65	325	4.8
69	Navy	United States	E4	Native American	1,983	82	307	15.5
70	Navy	United States	E4	Asian/Pacific Islander	2,903	43	133	4.6
71	Navy	United States	E4	Other	963	18	42	4.4
72	Navy	United States	E5-E6	White	66,195	830	1,973	3.0
73	Navy	United States	E5-E6	Black	24,631	298	897	3.6
74	Navy	United States	E5-E6	Hispanic	12,055	137	372	3.1
75	Navy	United States	E5-E6	Native American	2,154	117	255	11.8
76	Navy	United States	E5-E6	Asian/Pacific Islander	7,127	117	223	3.1
77	Navy	United States	E5-E6	Other	1,801	36	56	3.1
78	Navy	United States	E7-E9	White	20,042	244	401	2.0
79	Navy	United States	E7-E9	Black	4,487	56	108	2.4
80	Navy	United States	E7-E9	Hispanic	1,787	21	38	2.1
81	Navy	United States + Not United States	E7-E9	Native American	150	10	15	10.0
82	Navy	United States	E7-E9	Asian/Pacific Islander	1,510	26	37	2.5
83	Navy	United States	E7-E9	Other	875	18	22	2.5
84	Navy	United States	W1-O6	White	37,881	1,684	2,825	7.5

Table A-1.   (Continued)

85	Navy	United States	W1-O6	Black	3,538	306	655	18.5
86	Navy	United States	W1-O6	Hispanic	2,361	159	314	13.3
87	Navy	United States + Not United States	W1-O6	Native American	186	84	140	75.3
88	Navy	United States	W1-O6	Asian/Pacific Islander	1,367	141	212	15.5
89	Navy	United States	W1-O6	Other	1,248	134	172	13.8
90	Navy	Not United States	E1-E3	White	4,209	46	255	6.1
91	Navy	Not United States	E1-E3	Black	1,645	12	95	5.8
92	Navy	Not United States	E1-E3	Hispanic	727	59	419	57.6
93	Navy	Not United States	E1-E3	Native American	284	11	50	17.6
94	Navy	Not United States	E1-E3	Asian/Pacific Islander	474	37	136	28.7
95	Navy	Not United States	E1-E3	Other	295	5	13	4.4
96	Navy	Not United States	E4	White	3,433	48	158	4.6
97	Navy	Not United States	E4	Black	1,254	12	73	5.8
98	Navy	Not United States	E4	Hispanic	838	73	394	47.0
99	Navy	Not United States	E4	Native American	216	10	37	17.1
100	Navy	Not United States	E4	Asian/Pacific Islander	588	64	197	33.5
101	Navy	Not United States	E4	Other	138	3	7	5.1
102	Navy	Not United States	E5-E6	White	6,284	86	204	3.2
103	Navy	Not United States	E5-E6	Black	3,049	37	111	3.6
104	Navy	Not United States	E5-E6	Hispanic	1,438	160	403	28.0
105	Navy	Not United States	E5-E6	Native American	203	12	26	12.8
106	Navy	Not United States	E5-E6	Asian/Pacific Islander	1,733	265	504	29.1
107	Navy	Not United States	E5-E6	Other	303	6	9	3.0
108	Navy	Not United States	E7-E9	White	1,694	20	33	1.9
109	Navy	Not United States	E7-E9	Black	484	6	12	2.5

Table A-1. (Continued)

110	Navy	Not United States	E7-E9	Hispanic	195	28	50	25.6
111	Navy	Not United States	E7-E9	Asian/Pacific Islander	459	84	118	25.7
112	Navy	Not United States	E7-E9	Other	200	4	5	2.5
113	Navy	Not United States	W1-O6	White	3,562	160	286	8.0
114	Navy	Not United States	W1-O6	Black	433	40	86	19.9
115	Navy	Not United States	W1-O6	Hispanic	265	40	79	29.8
116	Navy	Not United States	W1-O6	Asian/Pacific Islander	244	54	81	33.2
117	Navy	Not United States	W1-O6	Other	180	22	28	15.6
118	Marine Corps	United States + Not United States	E1-E3	White	50,651	291	2,122	4.2
119	Marine Corps	United States + Not United States	E1-E3	Black	6,637	86	1,330	20.0
120	Marine Corps	United States + Not United States	E1-E3	Hispanic	9,629	316	3,704	38.5
121	Marine Corps	United States + Not United States	E1-E3	Native American	776	123	737	95.0
122	Marine Corps	United States + Not United States	E1-E3	Asian/Pacific Islander	1,815	138	848	46.7
123	Marine Corps	United States + Not United States	E1-E3	Other	3,934	32	243	6.2
124	Marine Corps	United States + Not United States	E4	White	15,408	98	578	3.8
125	Marine Corps	United States + Not United States	E4	Black	2,864	47	458	16.0
126	Marine Corps	United States + Not United States	E4	Hispanic	4,294	150	1,277	29.7
127	Marine Corps	United States + Not United States	E4	Native American	288	52	270	93.8
128	Marine Corps	United States + Not United States	E4	Asian/Pacific Islander	716	61	306	42.7

Table A-1.   (Continued)

129	Marine Corps	United States + Not United States	E4	Other	949	10	58	6.1
130	Marine Corps	United States + Not United States	E5-E6	White	20,759	177	560	2.7
131	Marine Corps	United States + Not United States	E5-E6	Black	6,422	152	699	10.9
132	Marine Corps	United States + Not United States	E5-E6	Hispanic	6,177	274	1,270	20.6
133	Marine Corps	United States + Not United States	E5-E6	Native American	376	93	309	82.2
134	Marine Corps	United States + Not United States	E5-E6	Asian/Pacific Islander	863	100	279	32.3
135	Marine Corps	United States + Not United States	E5-E6	Other	1,627	22	64	3.9
136	Marine Corps	United States + Not United States	E7-E9	White	7,074	67	136	1.9
137	Marine Corps	United States + Not United States	E7-E9	Black	3,211	100	270	8.4
138	Marine Corps	United States + Not United States	E7-E9	Hispanic	1,301	82	212	16.3
139	Marine Corps	United States + Not United States	E7-E9	Native American	103	33	69	67.0
140	Marine Corps	United States + Not United States	E7-E9	Asian/Pacific Islander	193	29	53	27.5
141	Marine Corps	United States + Not United States	E7-E9	Other	564	10	19	3.4
142	Marine Corps	United States + Not United States	W1-O6	White	14,008	926	2,086	14.9
143	Marine Corps	United States + Not United States	W1-O6	Black	1,219	116	409	33.6
144	Marine Corps	United States + Not United States	W1-O6	Hispanic	1,177	118	349	29.7
145	Marine Corps	United States + Not United States	W1-O6	Native American	102	50	97	95.1

Table A-1.   (Continued)

146	Marine Corps	United States + Not United States	W1-O6	Asian/Pacific Islander	339	56	114	33.6
147	Marine Corps	United States + Not United States	W1-O6	Other	1,248	113	235	18.8
148	Air Force	United States	E1-E3	White	51,824	883	2,240	4.3
149	Air Force	United States	E1-E3	Black	10,352	145	461	4.5
150	Air Force	United States	E1-E3	Hispanic	2,664	34	117	4.4
151	Air Force	United States + Not United States	E1-E3	Native American	518	91	293	56.6
152	Air Force	United States	E1-E3	Asian/Pacific Islander	2,539	82	218	8.6
153	Air Force	United States	E1-E3	Other	2,299	42	110	4.8
154	Air Force	United States	E4	White	27,128	459	1,126	4.2
155	Air Force	United States	E4	Black	7,587	116	338	4.5
156	Air Force	United States	E4	Hispanic	3,643	50	141	3.9
157	Air Force	United States + Not United States	E4	Native American	185	36	95	51.4
158	Air Force	United States	E4	Asian/Pacific Islander	1,025	36	81	7.9
159	Air Force	United States	E4	Other	2,166	45	99	4.6
160	Air Force	United States	E5-E6	White	67,681	1,257	2,114	3.1
161	Air Force	United States	E5-E6	Black	16,198	299	537	3.3
162	Air Force	United States	E5-E6	Hispanic	6,409	104	182	2.8
163	Air Force	United States	E5-E6	Native American	312	75	126	40.4
164	Air Force	United States	E5-E6	Asian/Pacific Islander	1,364	57	86	6.3
165	Air Force	United States	E5-E6	Other	3,913	91	137	3.5
166	Air Force	United States	E7-E9	White	20,638	347	462	2.2
167	Air Force	United States	E7-E9	Black	5,829	99	135	2.3
168	Air Force	United States	E7-E9	Hispanic	1,239	16	22	1.8
169	Air Force	United States + Not United States	E7-E9	Native American	71	20	27	38.0
170	Air Force	United States	E7-E9	Asian/Pacific Islander	280	13	16	5.7
171	Air Force	United States	E7-E9	Other	828	20	25	3.0
172	Air Force	United States	W1-O6	White	51,593	1,066	1,491	2.9
173	Air Force	United States	W1-O6	Black	4,142	330	478	11.5
174	Air Force	United States	W1-O6	Hispanic	2,250	170	244	10.8
175	Air Force	United States + Not United States	W1-O6	Native American	204	112	156	76.5

Table A-1.   (Continued)

176	Air Force	United States	W1-O6	Asian/Pacific Islander	1,266	134	173	13.7
177	Air Force	United States	W1-O6	Other	3,296	239	311	9.4
178	Air Force	Not United States	E1-E3	White	7,632	117	378	5.0
179	Air Force	Not United States	E1-E3	Black	1,835	25	88	4.8
180	Air Force	Not United States	E1-E3	Hispanic	581	53	183	31.5
181	Air Force	Not United States	E1-E3	Asian/Pacific Islander	301	60	159	52.8
182	Air Force	Not United States	E1-E3	Other	336	7	18	5.4
183	Air Force	Not United States	E4	White	6,608	107	284	4.3
184	Air Force	Not United States	E4	Black	2,183	32	93	4.3
185	Air Force	Not United States	E4	Hispanic	946	104	293	31.0
186	Air Force	Not United States	E4	Asian/Pacific Islander	245	48	107	43.7
187	Air Force	Not United States	E4	Other	561	12	26	4.6
188	Air Force	Not United States	E5-E6	White	15,787	291	489	3.1
189	Air Force	Not United States	E5-E6	Black	4,817	86	154	3.2
190	Air Force	Not United States	E5-E6	Hispanic	1,788	236	412	23.0
191	Air Force	Not United States	E5-E6	Native American	76	19	32	42.1
192	Air Force	Not United States	E5-E6	Asian/Pacific Islander	474	112	169	35.7
193	Air Force	Not United States	E5-E6	Other	1,293	30	45	3.5
194	Air Force	Not United States	E7-E9	White	4,395	67	89	2.0
195	Air Force	Not United States	E7-E9	Black	1,377	23	31	2.3
196	Air Force	Not United States	E7-E9	Hispanic	304	43	58	19.1
197	Air Force	Not United States	E7-E9	Asian/Pacific Islander	102	25	31	30.4
198	Air Force	Not United States	E7-E9	Other	302	7	9	3.0
199	Air Force	Not United States	W1-O6	White	6,691	141	197	2.9
200	Air Force	Not United States	W1-O6	Black	541	45	65	12.0

Table A-1.   (Continued)

201	Air Force	Not United States	W1-O6	Hispanic	326	53	76	23.3
202	Air Force	Not United States	W1-O6	Asian/Pacific Islander	219	61	79	36.1
203	Air Force	Not United States	W1-O6	Other	478	36	47	9.8
204	Coast Guard	United States + Not United States	E1-E3	White	4,516	86	278	6.2
205	Coast Guard	United States + Not United States	E1-E3	Black	371	44	155	41.8
206	Coast Guard	United States + Not United States	E1-E3	Hispanic	709	70	241	34.0
207	Coast Guard	United States + Not United States	E1-E3	Native American + Other	353	20	56	15.9
208	Coast Guard	United States + Not United States	E4	White	6,403	135	358	5.6
209	Coast Guard	United States + Not United States	E4	Black	321	42	122	38.0
210	Coast Guard	United States + Not United States	E4	Hispanic	622	68	192	30.9
211	Coast Guard	United States + Not United States	E4	Other	341	21	49	14.4
212	Coast Guard	United States + Not United States	E5-E6+E7-E9	White	13,162	355	560	4.3
213	Coast Guard	United States + Not United States	E5-E6+E7-E9	Black	1,087	185	308	28.3
214	Coast Guard	United States + Not United States	E5-E6+E7-E9	Hispanic	1,168	164	274	23.5
215	Coast Guard	United States + Not United States	E5-E6+E7-E9	Other	794	60	92	11.6
216	Coast Guard	United States + Not United States	W1-O6	White	6,091	175	245	4.0
217	Coast Guard	United States + Not United States	W1-O6	Black	379	70	101	26.6
218	Coast Guard	United States + Not United States	W1-O6	Hispanic	339	52	74	21.8

Table A-1.   (Continued)

219	Coast Guard	United States + Not United States	W1-O6	Other	328	27	36	11.0
220	Unknown	Unknown	Unknown	Unknown	2,337	60	147	6.3
Total					1,376,874	35,083	91,024	6.6

Table A-2.

Half-width Confidence Intervals, Precision Requirements, Domain Definitions, and Eligible Population Size by Domain for the 2005 WEOA

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
1	98.72%	1,374,537	0.5	0.035	All Domains	All Domains
2	96.05%	1,337,553	0.5	0.035	DoD	DoD
3	33.09%	462,572	0.5	0.035	Service	Army
4	25.23%	350,616	0.5	0.035		Navy
5	11.80%	164,724	0.5	0.035		Marine Corps
6	25.94%	359,641	0.5	0.035		Air Force
7	2.67%	36,984	0.5	0.035		Coast Guard
8	80.72%	1,122,259	0.5	0.035	Paygroup*DoD	E1 to E9*DoD
9	42.15%	584,871	0.5	0.035		E1 to E4*DoD
10	24.35%	337,115	0.5	0.035		E1 to E3*DoD
11	17.81%	247,756	0.5	0.035		E4*DoD
12	38.57%	537,388	0.5	0.035		E5 to E9*DoD
13	29.49%	409,051	0.5	0.035		E5 to E6*DoD
14	9.08%	128,337	0.5	0.035		E7 to E9*DoD
15	15.33%	215,294	0.5	0.035		*W1 to O6*DoD
16	1.07%	15,196	0.5	0.035		W1-W5*DoD
17	8.53%	119,874	0.5	0.035		O1-O3*DoD
18	5.72%	80,224	0.5	0.035		O4-O6*DoD
19	61.93%	860,793	0.5	0.035	Minority Status*DoD	non-Minority*DoD
20	34.12%	476,760	0.5	0.035		Minority*DoD
21	17.97%	249,993	0.5	0.035	Race/Ethnicity*DoD	non-Hispanic Black*DoD
22	8.75%	121,490	0.5	0.035		Hispanic (any race)*DoD
23	1.11%	15,413	0.5	0.035		Native American*DoD
24	3.24%	45,013	0.5	0.035		Asian & Pacific Islander*DoD
25	3.05%	44,851	0.5	0.035		Other & Unknown Race/Ethnicity*DoD

Table A-2.   (Continued)

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
26	81.84%	1,139,820	0.5	0.035	Gender*DoD	Male*DoD
27	14.21%	197,733	0.5	0.035		Female*DoD
28	79.04%	1,101,288	0.5		Region*DoD	US*DoD
29	8.69%	121,031	0.5		Detailed Region*DoD	North*DoD
30	45.63%	636,074	0.5			South*DoD
31	24.31%	338,398	0.5			West*DoD
32	17.01%	236,265	0.5		Region*DoD	Europe, Asia, Pacific Islands & Other*DoD
33	8.04%	111,559	0.5		Detailed Region*DoD	Europe*DoD
34	6.73%	93,449	0.5			Asia & Pacific Islander*DoD
35	2.24%	31,257	0.5		Region*DoD	Other Region*DoD
36	71.28%	990,515	0.5		Education*DoD	No College*DoD
37	6.84%	95,442	0.5			Some College*DoD
38	10.93%	153,395	0.5			4-year Degree*DoD
39	5.44%	76,259	0.5			Grad/Prof Degree*DoD
40	64.11%	890,965	0.5	0.035	All Domains*Minority Status	All Domains*non-Minority
41	61.93%	860,793	0.5	0.035	DoD*Minority Status	DoD*non-Minority
42	19.88%	277,281	0.5	0.035	Service*Minority Status	Army*non-Minority
43	15.53%	215,635	0.5	0.035		Navy*non-Minority
44	7.75%	107,900	0.5	0.035		Marine Corps*non-Minority
45	18.77%	259,977	0.5	0.035		Air Force*non-Minority
46	2.18%	30,172	0.5	0.036		Coast Guard*non-Minority
47	49.78%	690,707	0.5	0.035	Paygroup*Minority Status*DoD	E1 to E9*non-Minority*DoD
48	26.82%	371,692	0.5	0.035		E1 to E4*non-Minority*DoD
49	22.95%	319,015	0.5	0.035		E5 to E9*non-Minority*DoD
50	12.16%	170,086	0.5	0.035		W1 to O6*non-Minority*DoD
51	7.32%	102,467	0.5	0.035		W1 to O3*non-Minority*DoD
52	4.84%	67,619	0.5	0.035		W4 to O6*non-Minority*DoD
53	54.57%	758,582	0.5	0.035	Gender*Minority Status*DoD	Male*non-Minority*DoD
54	7.36%	102,211	0.5	0.035		Female*non-Minority*DoD

Table A-2.   (Continued)

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
55	51.55%	716,836	0.5		Region*Minority Status*DoD	US*non-Minority*DoD
56	6.21%	86,341	0.5		Detailed Region*Minority Status*DoD	North*non-Minority*DoD
57	29.15%	405,495	0.5			South*non-Minority*DoD
58	15.95%	221,615	0.5			West*non-Minority*DoD
59	10.38%	143,957	0.5		Region*Minority Status*DoD	Europe & Asia, Pacific Islands & Other*non-Minority*DoD
60	5.11%	70,878	0.5			Europe*non-Minority*DoD
61	3.84%	53,200	0.5			Asia & Pacific Islander*non-Minority*DoD
62	1.43%	19,879	0.5			Other Region*non-Minority*DoD
63	34.61%	483,572	0.5	0.035	All Domains*Minority Status	All Domains*Minority
64	34.12%	476,760	0.5	0.035	DoD*Minority Status	DoD*Minority
65	13.21%	185,291	0.5	0.035	Service*Minority Status	Army*Minority
66	9.70%	134,981	0.5	0.035		Navy*Minority
67	4.05%	56,824	0.5	0.035		Marine Corps*Minority
68	7.16%	99,664	0.5	0.035		Air Force*Minority
69	0.49%	6,812	0.5	0.035		Coast Guard*Minority
70	30.95%	431,552	0.5	0.035	Paygroup*Minority Status*DoD	E1 to E9*Minority*DoD
71	15.33%	213,179	0.5	0.035		E1 to E4*Minority*DoD
72	15.62%	218,373	0.5	0.035		E5 to E9*Minority*DoD
73	3.17%	45,208	0.5	0.035		W1 to O6*Minority*DoD
74	2.29%	32,603	0.5	0.035		W1 to O3*Minority*DoD
75	0.89%	12,605	0.5	0.035		W4 to O6*Minority*DoD
76	27.27%	381,238	0.5	0.035	Gender*Minority Status*DoD	Male*Minority*DoD
77	6.85%	95,522	0.5	0.035		Female*Minority*DoD
78	27.49%	384,452	0.5		Region*Minority Status*DoD	US*Minority*DoD
79	2.48%	34,690	0.5		Detailed Region*Minority Status*DoD	North*Minority*DoD
80	16.48%	230,579	0.5			South*Minority*DoD
81	8.36%	116,783	0.5			West*Minority*DoD
82	6.63%	92,308	0.5		Region*Minority Status*DoD	Europe & Asia, Pacific Islander & Other*Minority*DoD

Table A-2.   (Continued)

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
83	2.92%	40,681	0.5		Detailed Region*Minority Status*DoD	Europe*Minority*DoD
84	2.89%	40,249	0.5			Asia & Pacific Islands*Minority*DoD
85	0.81%	11,378	0.5		Region*Minority Status*DoD	Other Region*Minority*DoD
86	7.73%	107,936	0.5		Service*Race/Ethnicity	Army*non-Hispanic Black
87	3.40%	47,363	0.5			Army*Hispanic (any race)
88	0.27%	3,828	0.5	0.050		Army*Native American
89	0.96%	13,395	0.5	0.050		Army*Asian & Pacific Islander
90	4.82%	66,843	0.5	0.050		Navy*non-Hispanic Black
91	2.26%	31,399	0.5	0.050		Navy*Hispanic (any race)
92	0.62%	8,574	0.5	0.050		Navy*Native American
93	1.43%	19,877	0.5	0.050		Navy*Asian & Pacific Islander
94	1.46%	20,353	0.5	0.050		Marine Corps*non-Hispanic Black
95	1.62%	22,578	0.5	0.050		Marine Corps*Hispanic (any race)
96	0.12%	1,645	0.5	0.050		Marine Corps*Native American
97	0.28%	3,926	0.5	0.050		Marine Corps*Asian & Pacific Islander
98	3.96%	54,861	0.5	0.050		Air Force*non-Hispanic Black
99	1.46%	20,150	0.5	0.050		Air Force*Hispanic (any race)
100	0.10%	1,366	0.5	0.050		Air Force*Native American
101	0.56%	7,815	0.5	0.050		Air Force*Asian & Pacific Islander
102	0.16%	2,158	0.5	0.050		Coast Guard*non-Hispanic Black
103	0.21%	2,838	0.5	0.050		Coast Guard*Hispanic (any race)
104	0.05%	736	0.5			Coast Guard*Native American
105	16.57%	230,329	0.5		Paygroup*Race/Ethnicity*DoD	E1 to E9*non-Hispanic Black*DoD
106	8.02%	111,312	0.5			E1 to E9*Hispanic (any race)*DoD
107	1.05%	14,542	0.5	0.035		E1 to E9*Native American*DoD
108	2.83%	39,315	0.5	0.035		E1 to E9*Asian & Pacific Islander*DoD
109	1.40%	19,664	0.5			W1 to O6*non-Hispanic Black*DoD
110	0.73%	10,178	0.5			W1 to O6*Hispanic (any race)*DoD
111	0.06%	871	0.5	0.035		W1 to O6*Native American*DoD
112	0.41%	5,698	0.5	0.035		W1 to O6*Asian & Pacific Islander*DoD

Table A-2.   (Continued)

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
113	7.44%	103,176	0.5			E1 to E4*non-Hispanic Black*DoD
114	4.46%	61,792	0.5			E1 to E4*Hispanic (any race)*DoD
115	0.71%	9,765	0.5			E1 to E4*Native American*DoD
116	1.55%	21,508	0.5			E1 to E4*Asian & Pacific Islander*DoD
117	9.13%	127,153	0.5			E5 to E9*non-Hispanic Black*DoD
118	3.56%	49,520	0.5			E5 to E9*Hispanic (any race)*DoD
119	0.34%	4,777	0.5	0.050		E5 to E9*Native American*DoD
120	1.28%	17,807	0.5			E5 to E9*Asian & Pacific Islander*DoD
121	6.56%	91,748	0.5		Paygroup*Minority Status*DoD	O1-O3*non-Minority*DoD
122	1.97%	28,126	0.5			O1-O3*Minority*DoD
123	0.78%	10,996	0.5		Paygroup*Race/Ethnicity*DoD	O1-O3*non-Hispanic Black*DoD
124	0.47%	6,602	0.5			O1-O3*Hispanic (any race)*DoD
125	0.04%	528	0.5			O1-O3*Native American*DoD
126	0.29%	4,049	0.5			O1-O3*Asian & Pacific Islander*DoD
127	4.84%	67,619	0.5		Paygroup*Minority Status*DoD	O4-O6*non-Minority*DoD
128	0.89%	12,605	0.5			O4-O6*Minority*DoD
129	0.44%	6,133	0.5		Paygroup*Race/Ethnicity*DoD	O4-O6*non-Hispanic Black*DoD
130	0.19%	2,690	0.5			O4-O6*Hispanic (any race)*DoD
131	0.02%	252	0.5			O4-O6*Native American*DoD
132	0.10%	1,388	0.5			O4-O6*Asian & Pacific Islander*DoD
133	13.68%	190,355	0.5		Gender*Race/Ethnicity*DoD	Male*non-Hispanic Black*DoD
134	7.41%	102,923	0.5			Male*Hispanic (any race)*DoD
135	0.90%	12,518	0.5			Male*Native American*DoD
136	2.74%	38,147	0.5			Male*Asian & Pacific Islander*DoD
137	4.29%	59,638	0.5			Female*non-Hispanic Black*DoD
138	1.34%	18,567	0.5			Female*Hispanic (any race)*DoD
139	0.21%	2,895	0.5			Female*Native American*DoD
140	0.49%	6,866	0.5			Female*Asian & Pacific Islander*DoD
141	14.60%	203,282	0.5			US*non-Hispanic Black*DoD
142	6.98%	97,010	0.5			US*Hispanic (any race)*DoD

Table A-2.   (Continued)

Domain Number	Eligible Population Percentage	Domain Size	Prevalence	Interval Half Width	Domain	Domain Label
143	0.95%	13,171	0.5			US*Native American*DoD
144	2.55%	35,496	0.5			US*Asian & Pacific Islander*DoD
145	1.63%	22,619	0.5			Europe*non-Hispanic Black*DoD
146	0.77%	10,654	0.5	0.050		Europe*Hispanic (any race)*DoD
147	0.06%	886	0.5			Europe*Native American*DoD
148	0.20%	2,784	0.5	0.050		Europe*Asian & Pacific Islander*DoD
149	0.67%	9,264	0.5	0.050		Asia & Pacific Island*Hispanic (any race)*DoD
150	0.07%	1,006	0.5			Asia & Pacific Island*Native American*DoD
151	0.42%	5,886	0.5	0.050		Asia & Pacific Island*Asian & Pacific Islander*DoD

1. The domain sizes exclude 2,337 persons classified into the unknown stratum.
2. The precision constraint is given as the maximum half-width of a 95% confidence interval.

Appendix B.
Nonresponse Adjustment Cell Definitions

Table B-1.
Nonresponse Adjustment Cell Definitions and Adjustment Factors

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
1	1, 7	Service: Army Race-Ethnicity: White Paygrade: E1-E4 Location: In US	5.4374	1.1994
2	2, 8	Service: Army Race-Ethnicity: Black Paygrade: E1-E4 Location: In US	6.2196	1.3112
3	3, 9	Service: Army Race-Ethnicity: Hispanic Paygrade: E1-E4 Location: In US	6.2484	1.3544
4	4, 10, 34, 40	Service: Army Race-Ethnicity: American Indian/Alaskan Native Paygrade: E1-E4 Location: All	5.6903	1.1722
5	5	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E3 Location: In US	4.2162	1.1935
6	6, 12, 18, 36, 42, 48	Service: Army Race-Ethnicity: Other Paygrade: E1-E6 Location: All	2.5940	1.1550
7	11	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E4 Location: In US	3.2955	1.2571
8	13	Service: Army Race-Ethnicity: White Paygrade: E5-E6 Location: In US	2.3974	1.1292
9	14	Service: Army Race-Ethnicity: Black Paygrade: E5-E6 Location: In US	2.4524	1.1467
10	15	Service: Army Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: In US	2.3300	1.1512

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
11	16, 46	Service: Army Race-Ethnicity: American Indian/Alaskan Native Paygrade: E5-E6 Location: All	2.1837	1.1530
12	17	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E5-E6 Location: In US	2.4444	1.1290
13	19	Service: Army Race-Ethnicity: White Paygrade: E7-E9 Location: In US	1.5794	1.1042
14	20	Service: Army Race-Ethnicity: Black Paygrade: E7-E9 Location: In US	1.6435	1.1622
15	21, 51	Service: Army Race-Ethnicity: Hispanic Paygrade: E7-E9 Location: All	1.5682	1.0756
16	22, 52	Service: Army, Army Race-Ethnicity: American Indian/Alaskan Paygrade: E7-E9 Location: All	1.3994	1.1005
17	23, 53	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E7-E9 Location: All	1.6250	1.0137
18	24, 54	Service: Army Race-Ethnicity: Other Paygrade: E7-E9 Location: All	1.6530	1.0954
19	25	Service: Army Race-Ethnicity: White Paygrade: W1-W2 Location: In US Months away on active duty service: 0.321-3.34 months	1.4310	1.0755
20	25	Service: Army Race-Ethnicity: White Paygrade: W1-W2 Location: In US Months away on active duty service: 3.35-4.86 months, unknown	2.5473	1.1053

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
21	25	Service: Army Race-Ethnicity: White Paygrade: W3-W5 Location: In US Months away on active duty service: 0.321-1.83-2.58 months, unknown	1.3095	1.0921
22	25	Service: Army Race-Ethnicity: White Paygrade: W3-W5 Location: In US Age: 42 years and younger Months away on active duty service: 2.59-4.86 months	2.1000	1.0693
23	25	Service: Army Race-Ethnicity: White Paygrade: W3-W5 Location: In US Age: 43 years and older Months away on active duty service: 2.59-4.86 months	1.4429	1.1311
24	25	Service: Army Race-Ethnicity: White Paygrade: O1 Location: In US Gender: Male	2.5556	1.2073
25	25	Service: Army Race-Ethnicity: White Paygrade: O1 Location: In US Gender: Female	1.9434	1.0600
26	25	Service: Army Race-Ethnicity: White Paygrade: O2 Location: In US Family Status: Married or single, with children	1.7308	1.1183
27	25	Service: Army Race-Ethnicity: White Paygrade: O2 Location: In US Gender: Male Family Status: Married or single, without children	2.3652	1.2000
28	25	Service: Army Race-Ethnicity: White Paygrade: O2 Location: In US Gender: Female Family Status: Married or single, without children	1.8393	1.1277

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
29	25	Service: Army Race-Ethnicity: White Paygrade: O3 Location: In US Family Status: Married without children, single with children	1.6060	1.0772
30	25	Service: Army Race-Ethnicity: White Paygrade: O3 Location: In US Family Status: Single without children	1.7917	1.1215
31	25	Service: Army Race-Ethnicity: White Paygrade: O3 Location: In US Age: 30 years and younger Family Status: Married with children	1.6875	1.0596
32	25	Service: Army Race-Ethnicity: White Paygrade: O3 Location: In US Age: 52 years and older Family Status: Married with children	1.3652	1.1010
33	25	Service: Army Race-Ethnicity: White Paygrade: O4 Location: In US	1.2992	1.0710
34	25	Service: Army Race-Ethnicity: White Paygrade: O5-O6 Location: In US	1.2331	1.0409
35	26	Service: Army Race-Ethnicity: Black Paygrade: All Officers Location: In US Age: 30 years and younger	2.3929	1.1429
36	26	Service: Army Race-Ethnicity: Black Paygrade: All Officers Location: In US Age: 31-38 years	1.5765	1.1108
37	26	Service: Army Race-Ethnicity: Black Paygrade: All Officers Location: In US Age: 39 years and older Gender: Male	1.3803	1.0903

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
38	26	Service: Army Race-Ethnicity: Black Paygrade: All Officers Location: In US Age: 39 years and older Gender: Female	1.5979	1.0899
39	27	Service: Army Race-Ethnicity: Hispanic Paygrade: All Officers Location: In US	1.5270	1.1012
40	28	Service: Army Race-Ethnicity: American Indian/Alaskan Native Paygrade: All Officers Location: Not in US	1.5497	1.0814
41	29	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: In US	1.6348	1.1210
42	30	Service: Army Race-Ethnicity: Other Paygrade: All Officers Location: In US	1.7149	1.1667
43	31, 37	Service: Army Race-Ethnicity: White Paygrade: E1-E4 Location: Not in US	4.9868	1.1794
44	32, 38	Service: Army Race-Ethnicity: Black Paygrade: E1-E4 Location: Not in US	5.1453	1.2668
45	33, 39	Service: Army, Race-Ethnicity: Hispanic Paygrade: E1-E4 Location: Not in US	4.5391	1.2604
46	35	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E3 Location: Not in US	3.4437	1.2931
47	41	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E4 Location: Not in US	3.1914	1.2778
48	43	Service: Army Race-Ethnicity: White Paygrade: E5-E6 Location: Not in US	2.1688	1.1357

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
49	44	Service: Army Race-Ethnicity: Black Paygrade: E5-E6 Location: Not in US	2.3769	1.2056
50	45	Service: Army Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: Not in US	2.2996	1.1558
51	47	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: E5-E6 Location: Not in US	1.9205	1.1986
52	49	Service: Army Race-Ethnicity: White Paygrade: E7-E9 Location: Not in US	1.5870	1.0455
53	50	Service: Army Race-Ethnicity: Black Paygrade: E7-E9 Location: Not in US	1.7581	1.0333
54	55	Service: Army Race-Ethnicity: White Paygrade: W1-W2 Location: Not in US	1.8511	1.0682
55	55	Service: Army Race-Ethnicity: White Paygrade: W3-W5 Location: Not in US	1.4455	1.0989
56	55	Service: Army Race-Ethnicity: White Paygrade: O1-O2 Location: Not in US	2.3291	1.1135
57	55	Service: Army Race-Ethnicity: White Paygrade: O3 Location: Not in US	1.5675	1.0946
58	55	Service: Army Race-Ethnicity: White Paygrade: O4-O6 Location: Not in US	1.2949	1.0567
59	56	Service: Army Race-Ethnicity: Black Paygrade: All Officers Location: Not in US	1.8583	1.1065

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
60	57	Service: Army Race-Ethnicity: Hispanic Paygrade: All Officers Location: Not in US	1.6991	1.1429
61	58	Service: Army Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: Not in US	1.5745	1.0938
62	59	Service: Army Race-Ethnicity: Other Paygrade: All Officers Location: Not in US	1.7273	1.1000
63	60, 66	Service: Navy Race-Ethnicity: White Paygrade: E1-E4 Location: In US	3.6570	1.1614
64	61, 67, 91, 97	Service: Navy Race-Ethnicity: Black Paygrade: E1-E4 Location: All	4.9251	1.3226
65	62, 68	Service: Navy Race-Ethnicity: Hispanic Paygrade: E1-E4 Location: In US	3.0462	1.2424
66	63, 69, 93, 99	Service: Navy Race-Ethnicity: American Indian/Alaskan Paygrade: E1- E4 Location: All	4.2142	1.1594
67	64, 70	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E4 Location: In US	3.3555	1.1638
68	65, 71, 95, 101	Service: Navy Race-Ethnicity: Other Paygrade: E1-E4 Location: All	2.9670	1.2324
69	72	Service: Navy Race-Ethnicity: White Paygrade: E5 Location: In US Marital Status: Married	2.5989	1.0809
70	72	Service: Navy Race-Ethnicity: White Paygrade: E5 Location: In US Marital Status: Not Married	2.1600	1.1062

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
71	72	Service: Navy Race-Ethnicity: White Paygrade: E6 Location: In US	1.7809	1.0696
72	73	Service: Navy Race-Ethnicity: Black Paygrade: E5-E6 Location: In US	2.4171	1.1173
73	74	Service: Navy Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: In US	2.1420	1.0563
74	75, 81, 105	Service: Navy Race-Ethnicity: American Indian/Alaskan Native Paygrade: E5-E9 Location: All	2.0737	1.0782
75	76	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E5-E6 Location: In US	1.9640	1.1327
76	77, 83, 107, 112	Service: Navy Race-Ethnicity: Other, Paygrade: E5-E9 Location: All	1.6685	1.0000
77	78, 108	Service: Navy Race-Ethnicity: White Paygrade: E7-E9 Location: All	1.4027	1.0318
78	79, 109	Service: Navy Race-Ethnicity: Black Paygrade: E7-E9 Location: All	1.5524	1.0278
79	80, 110	Service: Navy Race-Ethnicity: Hispanic Paygrade: E7-E9 Location: All	1.3809	1.0028
80	82, 111	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E7-E9 Location: All	1.2289	1.0374
81	84	Service: Navy Race-Ethnicity: White Paygrade: W1-W5 Location: In US	1.3529	1.0200

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
82	84	Service: Navy Race-Ethnicity: White Paygrade: O1-O2 Location: In US Months away on active duty service: 0.321-1.82 months	1.3810	1.0328
83	84	Service: Navy Race-Ethnicity: White Paygrade: O1-O2 Location: In US Months away on active duty service: 1.83-4.86 months, unknown Family Status: Married without children, single with children	1.5281	1.0471
84	84	Service: Navy Race-Ethnicity: White Paygrade: O1-O2 Location: In US Months away on active duty service: 1.83-4.86 months, unknown Family Status: Married with children, single without children	2.0356	1.0331
85	84	Service: Navy Race-Ethnicity: White Paygrade: O3 Location: In US Marital Status: Married	1.6289	1.0324
86	84	Service: Navy Race-Ethnicity: White Paygrade: O3 Location: In US Marital Status: Not Married	1.4026	1.0542
87	84	Service: Navy Race-Ethnicity: White Paygrade: O4-O6 Location: In US Age: 38 years and younger Marital Status: Married	1.7500	1.0588
88	84	Service: Navy Race-Ethnicity: White Paygrade: O4-O6 Location: In US Age: 62 years and older Marital Status: Married	1.3333	1.0615

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
89	84	Service: Navy Race-Ethnicity: White Paygrade: O4-O6 Location: In US Age: 31 years and younger Marital Status: Not Married	1.5410	1.0517
90	84	Service: Navy Race-Ethnicity: White Paygrade: O4-O6 Location: In US Age: 35 years and older Marital Status: Not Married	1.2653	1.0483
91	85	Service: Navy Race-Ethnicity: Black Paygrade: All Officers Location: In US	1.8271	1.0743
92	86	Service: Navy Race-Ethnicity: Hispanic Paygrade: All Officers Location: In US	1.6667	1.0702
93	87	Service: Navy Race-Ethnicity: American Indian/Alaskan Native Paygrade: All Officers Location: Not in US	1.6353	1.0759
94	88	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: In US	1.4545	1.1172
95	89, 117	Service: Navy Race-Ethnicity: Other Paygrade: All Officers Location: All	1.6688	1.0266
96	90	Service: Navy Race-Ethnicity: White Paygrade: E1-E3 Location: Not in US	3.4714	1.2069
97	92	Service: Navy Race-Ethnicity: Hispanic Paygrade: E1-E3 Location: Not in US	3.3109	1.1346
98	94	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E3 Location: Not in US	2.7872	1.2368

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
99	96	Service: Navy Race-Ethnicity: White Paygrade: E4 Location: Not in US	3.2609	1.1795
100	98	Service: Navy Race-Ethnicity: Hispanic Paygrade: E4 Location: Not in US	2.7939	1.1909
101	100	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E4 Location: Not in US	2.1954	1.0741
102	102	Service: Navy Race-Ethnicity: White Paygrade: E5-E6 Location: Not in US	1.7345	1.0463
103	103	Service: Navy Race-Ethnicity: Black Paygrade: E5-E6 Location: Not in US	2.4222	1.1579
104	104	Service: Navy Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: Not in US	1.9406	1.0924
105	106	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: E5-E6 Location: Not in US	1.7079	1.0940
106	113	Service: Navy Race-Ethnicity: White Paygrade: All Officers Location: Not in US	1.3930	1.0806
107	114	Service: Navy Race-Ethnicity: Black Paygrade: All Officers Location: Not in US	1.8085	1.0217
108	115	Service: Navy Race-Ethnicity: Hispanic Paygrade: All Officers Location: Not in US	1.9024	1.0513
109	116	Service: Navy Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: Not in US	1.3621	1.0741

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
110	118, 124	Service: Marine Corps Race-Ethnicity: White Paygrade: E1- E4 Location: Not in US	6.6902	1.2138
111	119, 125	Service: Marine Corps Race-Ethnicity: Black Paygrade: E1-E4 Location: Not in US	7.2334	1.3154
112	120, 126	Service: Marine Corps Race-Ethnicity: Hispanic Paygrade: E1-E4 Location: Not in US	6.7195	1.2351
113	121, 127	Service: Marine Corps Race-Ethnicity: American Indian/Alaskan Native Paygrade: E1-E4 Location: Not in US	6.8287	1.2427
114	122, 128	Service: Marine Corps Race-Ethnicity: Asian, Pacific Islander Paygrade: E1- E4 Location: Not in US	4.7148	1.2311
115	123, 129, 135, 141	Service: Marine Corps Race-Ethnicity: Other Paygrade: E1-E9 Location: Not in US	5.1256	1.1978
116	130	Service: Marine Corps Race-Ethnicity: White Paygrade: E5-E6 Location: Not in US	3.0345	1.0741
117	131	Service: Marine Corps Race-Ethnicity: Black Paygrade: E5-E6 Location: Not in US	3.0360	1.1632
118	132	Service: Marine Corps Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: Not in US	3.0574	1.1335
119	133	Service: Marine Corps Race-Ethnicity: American Indian/Alaskan Native Paygrade: E5-E6 Location: Not in US	2.8641	1.1705
120	134, 140	Service: Marine Corps Race-Ethnicity: Asian, Pacific Islander Paygrade: E5- E9 Location: Not in US	2.5617	1.0447

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
121	136	Service: Marine Corps Race-Ethnicity: White Paygrade: E7-E9 Location: Not in US	1.7013	1.0417
122	137	Service: Marine Corps Race-Ethnicity: Black Paygrade: E7-E9 Location: Not in US	1.9412	1.1356
123	138	Service: Marine Corps Race-Ethnicity: Hispanic Paygrade: E7-E9 Location: Not in US	2.0000	1.1075
124	139	Service: Marine Corps Race-Ethnicity: American Indian/Alaskan Native Paygrade: E7-E9 Location: Not in US	1.9706	1.1333
125	142	Service: Marine Corps Race-Ethnicity: White Paygrade: All Officers Location: Not in US Age: 26 years and younger	2.5733	1.0797
126	142	Service: Marine Corps Race-Ethnicity: White Paygrade: All Officers Location: Not in US Age: 27-32 years	2.1384	1.1206
127	142	Service: Marine Corps Race-Ethnicity: White Paygrade: All Officers Location: Not in US Age: 33-36 years	1.7782	1.1070
128	142	Service: Marine Corps Race-Ethnicity: White Paygrade: All Officers Location: Not in US Age: 37 years and older	1.3876	1.1071
129	143	Service: Marine Corps Race-Ethnicity: Black Paygrade: All Officers Location: Not in US	2.2123	1.0988
130	144	Service: Marine Corps Race-Ethnicity: Hispanic Paygrade: All Officers Location: Not in US	2.0602	1.0573

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
131	145	Service: Marine Corps Race-Ethnicity: American Indian/Alaskan Native Paygrade: All Officers Location: Not in US	2.0652	1.2105
132	146	Service: Marine Corps Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: Not in US	1.8065	1.1321
133	147	Service: Marine Corps Race-Ethnicity: Other Paygrade: All Officers Location: Not in US	2.2308	1.0833
134	148	Service: Air Force Race-Ethnicity: White Paygrade: E1-E3 Location: In US Gender: Male	2.2859	1.1296
135	148	Service: Air Force Race-Ethnicity: White Paygrade: E1-E3 Location: In US Gender: Female	1.8908	1.1075
136	149, 179	Service: Air Force Race-Ethnicity: Black Paygrade: E1-E3 Location: All	2.8350	1.2350
137	150	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E1-E3 Location: In US	2.1176	1.0625
138	151	Service: Air Force Race-Ethnicity: American Indian/Alaskan Native Paygrade: E1-E3 Location: Not in US	2.5182	1.1224
139	152	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E3 Location: In US	2.1443	1.1149
140	153, 182	Service: Air Force Race-Ethnicity: Other Paygrade: E1-E3 Location: All	2.4864	1.0660
141	154	Service: Air Force Race-Ethnicity: White Paygrade: E4 Location: In US	2.2607	1.0939

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
142	155	Service: Air Force Race-Ethnicity: Black Paygrade: E4 Location: In US	2.7241	1.1616
143	156	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E4 Location: In US	2.4815	1.1042
144	157	Service: Air Force Race-Ethnicity: American Indian/Alaskan Native Paygrade: E4 Location: Not in US	2.5882	1.1333
145	158	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E4 Location: In US	2.2286	1.1667
146	159, 187	Service: Air Force Race-Ethnicity: Other Paygrade: E4 Location: All	2.4465	1.2897
147	160	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: In US Age: 32 years and younger Dual Spouse: Not Dual Service spouse	1.8149	1.0827
148	160	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: In US Age: 32 years and younger Dual Spouse: Dual Active/Reserve	2.2065	1.0952
149	160	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: In US Age: 33-40 years Marital Status: Married	1.7821	1.0400
150	160	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: In US Age: 33-40 years Marital Status: Not Married	1.4793	1.0669

Table B-1. **(Continued)**

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
151	160	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: In US Age: 41 years and older	1.2687	1.0469
152	161	Service: Air Force Race-Ethnicity: Black Paygrade: E5-E6 Location: In US	1.9407	1.1538
153	162	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: In US	1.6574	1.0693
154	163, 169, 191	Service: Air Force Race-Ethnicity: American Indian/Alaskan Native Paygrade: E5-E9 Location: All	1.7687	1.0984
155	164, 170, 197	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E5- E9 Location: All	1.6844	1.1315
156	165, 171, 193, 198	Service: Air Force Race-Ethnicity: Other Paygrade: E5-E9 Location: All	1.7673	1.0367
157	166	Service: Air Force Race-Ethnicity: White Paygrade: E7-E9 Location: In US	1.3458	1.0425
158	167, 195	Service: Air Force Race-Ethnicity: Black Paygrade: E7-E9 Location: All	1.4812	1.0758
159	168, 196	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E7-E9 Location: All	1.5622	1.0116
160	172	Service: Air Force Race-Ethnicity: White Paygrade: All Officers Location: In US Marital Status: Married	1.6891	1.0485

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment $(f_c{}^{A1})$	Eligible Nonresponse Adjustment $(f_c{}^{A2})$
161	172	Service: Air Force Race-Ethnicity: White Paygrade: W1-W5, O1-O4 Location: In US Months away on active duty service: 0.321-1.06 months Marital Status: Not Married	1.3798	1.0574
162	172	Service: Air Force Race-Ethnicity: White Paygrade: W1-W5, O1-O4 Location: In US Months away on active duty service: 1.07-1.82 months Marital Status: Not Married	1.7925	1.0400
163	172	Service: Air Force Race-Ethnicity: White Paygrade: W1-W5, O1-O4 Location: In US Months away on active duty service: 1.83-2.58 months Marital Status: Not Married	1.4204	1.0426
164	172	Service: Air Force Race-Ethnicity: White Paygrade: W1-W5, O1-O4 Location: In US Months away on active duty service: 2.59-4.10 months, unknown Marital Status: Not Married	1.6600	1.0417
165	172	Service: Air Force Race-Ethnicity: White Paygrade: O5-O6 Location: In US Marital Status: Not Married	1.2287	1.0474
166	173	Service: Air Force Race-Ethnicity: Black Paygrade: All Officers Location: In US	1.7992	1.0661
167	174	Service: Air Force Race-Ethnicity: Hispanic Paygrade: All Officers Location: In US	1.3851	1.0359
168	175	Service: Air Force Race-Ethnicity: American Indian/Alaskan Native Paygrade: All Officers Location: Not in US	1.5876	1.0659

Table B-1.   (Continued)

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
169	176	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: In US	1.4250	1.0619
170	177, 203	Service: Air Force Race-Ethnicity: Other Paygrade: All Officers Location: All	1.5667	1.0466
171	178	Service: Air Force Race-Ethnicity: White Paygrade: E1-E3 Location: Not in US	2.4865	1.1855
172	180	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E1-E3 Location: Not in US	2.1585	1.1389
173	181	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E1-E3 Location: Not in US	2.2714	1.1475
174	183	Service: Air Force Race-Ethnicity: White Paygrade: E4 Location: Not in US	2.9043	1.1059
175	184	Service: Air Force Race-Ethnicity: Black Paygrade: E4 Location: Not in US	2.3333	1.1818
176	185	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E4 Location: Not in US	2.3898	1.1346
177	186	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E4 Location: Not in US	2.3261	1.1220
178	188	Service: Air Force Race-Ethnicity: White Paygrade: E5-E6 Location: Not in US	1.7732	1.0593
179	189	Service: Air Force Race-Ethnicity: Black Paygrade: E5-E6 Location: Not in US	2.0959	1.0735

Table B-1. **(Continued)**

Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
180	190	Service: Air Force Race-Ethnicity: Hispanic Paygrade: E5-E6 Location: Not in US	1.9023	1.1026
181	192	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: E5-E6 Location: Not in US	1.7320	1.0319
182	194	Service: Air Force Race-Ethnicity: White Paygrade: E7-E9 Location: Not in US	1.4032	1.0517
183	199	Service: Air Force Race-Ethnicity: White Paygrade: All Officers Location: Not in US	1.6281	1.0342
184	200	Service: Air Force Race-Ethnicity: Black Paygrade: All Officers Location: Not in US	1.6842	1.0556
185	201	Service: Air Force Race-Ethnicity: Hispanic Paygrade: All Officers Location: Not in US	1.5000	1.0638
186	202	Service: Air Force Race-Ethnicity: Asian, Pacific Islander Paygrade: All Officers Location: Not in US	1.7727	1.0732
187	204, 208	Service: Coast Guard Race-Ethnicity: White Paygrade: E1- E4 Location: In US	2.8487	1.1038
188	205, 209	Service: Coast Guard Race-Ethnicity: Black Paygrade: E1-E4 Location: In US	3.5227	1.1691
189	206, 210	Service: Coast Guard Race-Ethnicity: Hispanic Paygrade: E1-E4 Location: In US	2.8731	1.0882
190	207, 211, 215, 219	Service: Coast Guard Race-Ethnicity: Other Paygrade: E1- E9, All Officers Location: All	1.8936	1.0963

Table B-1.   (Continued)

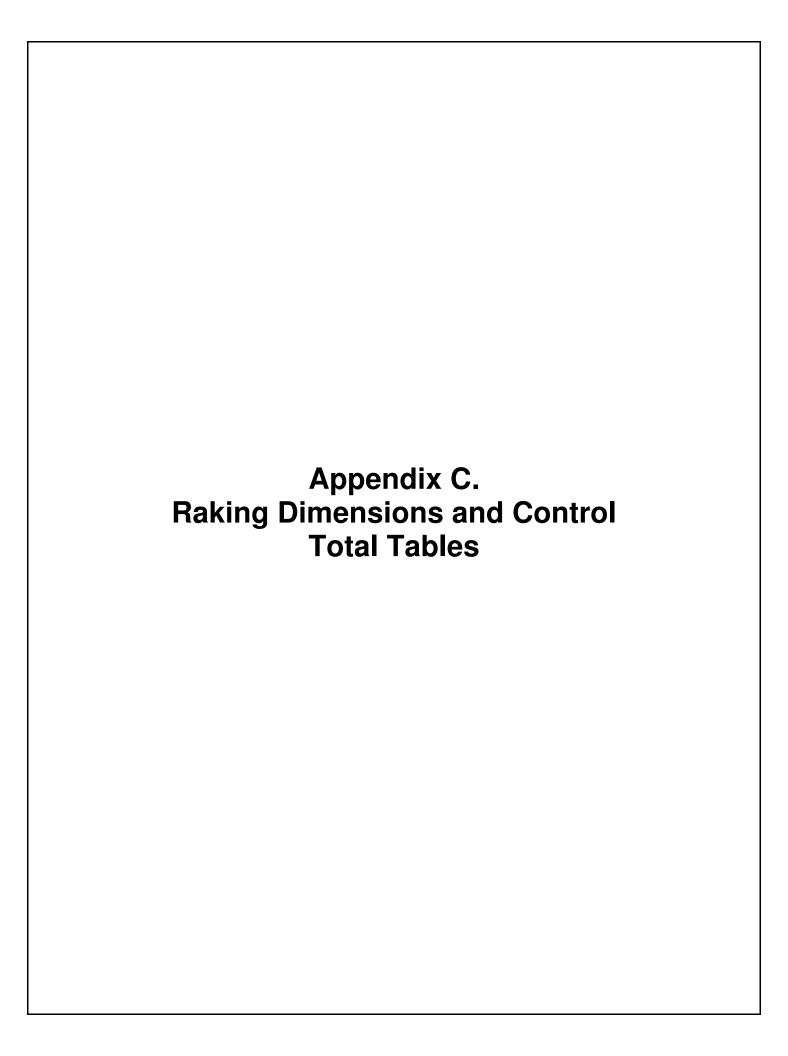| Weighting Class	Stratum	Description	Unknown Eligibility Adjustment (f_c^{A1})	Eligible Nonresponse Adjustment (f_c^{A2})
191	212	Service: Coast Guard Race-Ethnicity: White Paygrade: E7-E9 Location: In US	1.7419	1.0302
192	213	Service: Coast Guard Race-Ethnicity: Black Paygrade: E7-E9 Location: In US	2.1197	1.0853
193	214	Service: Coast Guard Race-Ethnicity: Hispanic Paygrade: E7-E9 Location: In US	2.0000	1.0894
194	216	Service: Coast Guard Race-Ethnicity: White Paygrade: All Officers Location: In US	1.3005	1.0225
195	217	Service: Coast Guard Race-Ethnicity: Black Paygrade: All Officers Location: In US	1.3333	1.0571
196	218	Service: Coast Guard Race-Ethnicity: Hispanic Paygrade: All Officers Location: In US	1.4231	1.0400
197	220	Service: Marine Corps Race-Ethnicity: Black Paygrade: All Officers Location: unknown	2.4211	1.0556

**Appendix C.
Raking Dimensions and Control
Total Tables**

Table C-1.

Definition and Control Total of the Dimension DIM1 Used in Raking

DIM1	Service Branch	Gender	Age Category	Control Total
111	Army	Male	17 to 24 years, Unknown	136,394
112	Army	Male	25 to 29 years	86,174
113	Army	Male	30 to 34 years	63,112
114	Army	Male	35 to 39 years	51,517
115	Army	Male	40 to 44 years	28,772
116	Army	Male	45 to 49 years	9,811
117	Army	Male	50 years and older	3,329
121	Army	Female	17 to 24 years, Unknown	25,324
122	Army	Female	25 to 29 years	14,207
123	Army	Female	30 to 34 years	9,791
124	Army	Female	35 to 39 years	7,352
125	Army	Female	40 to 44 years	4,218
126	Army	Female	45 to 49 years	1,603
127	Army	Female	50 years and older	689
211	Navy	Male	17 to 24 years, Unknown	101,160
212	Navy	Male	25 to 29 years	63,659
213	Navy	Male	30 to 34 years	46,726
214	Navy	Male	35 to 39 years	41,814
215	Navy	Male	40 to 44 years	24,008
216	Navy	Male	45 to 49 years	8,572
217	Navy	Male	50 years and older	2,852
221	Navy	Female	17 to 24 years, Unknown	21,559
222	Navy	Female	25 to 29 years	11,323
223	Navy	Female	30 to 34 years	5,799
224	Navy	Female	35 to 39 years	4,678
225	Navy	Female	40 to 44 years	2,893
226	Navy	Female	45 years and older	1,643
311	Air Force	Male	17 to 24 years, Unknown	83,652
312	Air Force	Male	25 to 29 years	27,732
313	Air Force	Male	30 to 34 years	16,192
314	Air Force	Male	35 to 39 years	11,125
315	Air Force	Male	40 to 44 years	5,673
316	Air Force	Male	45 years and older	2,372
321	Air Force	Female	17 to 24 years, Unknown	5,825
322	Air Force	Female	25 to 29 years	1,903
323	Air Force	Female	30 years and older	1,788

Table C-1.   (Continued)

DIM1	Service Branch	Gender	Age Category	Control Total
411	Marine Corps	Male	17 to 24 years, Unknown	87,291
412	Marine Corps	Male	25 to 29 years	61,290
413	Marine Corps	Male	30 to 34 years	43,809
414	Marine Corps	Male	35 to 39 years	44,121
415	Marine Corps	Male	40 to 44 years	31,739
416	Marine Corps	Male	45 to 49 years	9,272
417	Marine Corps	Male	50 years and older	2,406
421	Marine Corps	Female	17 to 24 years, Unknown	26,919
422	Marine Corps	Female	25 to 29 years	17,004
423	Marine Corps	Female	30 to 34 years	9,626
424	Marine Corps	Female	35 to 39 years	7,332
425	Marine Corps	Female	40 to 44 years	4,726
426	Marine Corps	Female	45 years and older	2,212
511	Coast Guard	Male	17 to 24 years, Unknown	9,537
512	Coast Guard	Male	25 to 29 years	7,528
513	Coast Guard	Male	30 to 34 years	5,210
514	Coast Guard	Male	35 to 39 years	4,365
515	Coast Guard	Male	40 to 44 years	3,852
516	Coast Guard	Male	45 years and older	1,976
521	Coast Guard	Female	17 to 29 years, Unknown	2,586
523	Coast Guard	Female	30 years and older	1,366
Total				1,319,408

Table C-2.
Definition and Control Total of the Dimension DIM2 Used in Raking

DIM2	Service Branch	Paygrade Group	Control Total
11	Army	E1 to E4, Unknown Enlisted	182,127
12	Army	E5 to E9	186,258
13	Army	W1 to W5	11,859
14	Army	O1 to O3	35,391
15	Army	O4 to O6	26,658
21	Navy	E1 to E4	124,041
22	Navy	E5 to E9	162,429
23	Navy	W1 to W5	1,628
24	Navy	O1 to O3	28,146
25	Navy	O4 to O6	20,442
31	Marine Corps	E1 to E4	86,980
32	Marine Corps	E5 to E9	51,372
33	Marine Corps	W1 to W5	1,847
34	Marine Corps	O1 to O3	10,144
35	Marine Corps	O4 to O6	5,919
41	Air Force	E1 to E4	118,790
42	Air Force	E5 to E9	158,647
44	Air Force	O1 to O3	40,781
45	Air Force	O4 to O6	29,529
51	Coast Guard	E1 to E4	12,338
52	Coast Guard	E5 to E9	16,888
53	Coast Guard	W1 to W5	1,468
54	Coast Guard	O1 to O3	3,470
55	Coast Guard	O4 to O6	2,256
Total			1,319,408

Table C-3.
Definition and Control Total of the Dimension DIM3 Used in Raking

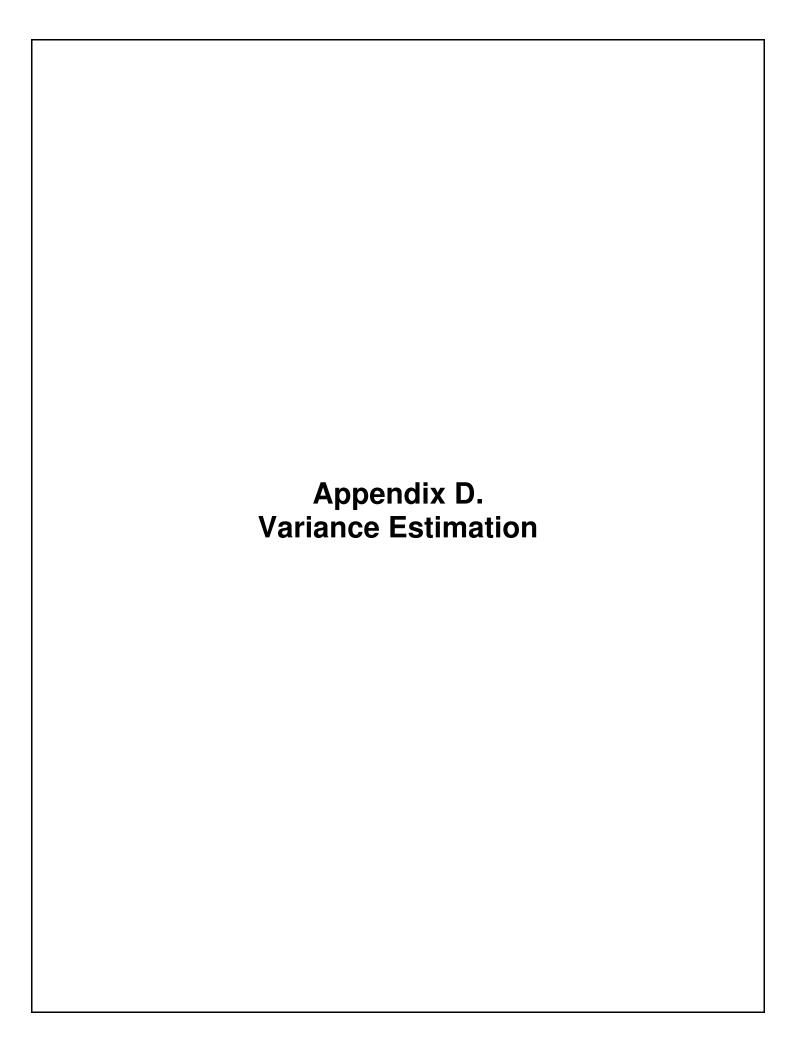| DIM3	Service Branch	Race/Ethnicity	Control Total
11	Army	White	264,613
12	Army	Black	103,366
13	Army	Hispanic	45,652
14	Army	Asian	12,953
16	Army	American Indian/Alaskan Native	3,667
18	Army	Unknown	12,042
21	Navy	White	206,506
22	Navy	Black	64,119
23	Navy	Hispanic	30,378
24	Navy	Asian, Hawaiian/Pacific Islander	19,402
26	Navy	American Indian/Alaskan Native	8,220
27	Navy	Multi-race, Unknown	8,061
31	Marine	White	102,302
32	Marine	Black	19,370
33	Marine	Hispanic	22,195
34	Marine	Asian	3,154
35	Marine	Hawaiian/Pacific Islander	698
36	Marine	American Indian/Alaskan Native	1,622
37	Marine	Multi-race, Unknown	6,921
41	Air Force	White	250,696
42	Air Force	Black	52,863
43	Air Force	Hispanic	20,324
44	Air Force	Asian	6,641
45	Air Force	Hawaiian/Pacific Islander	1,423
46	Air Force	American Indian/Alaskan Native	1,382
47	Air Force	Multi-race	3,557
48	Air Force	Unknown	10,861
51	Coast Guard	White	29,369
52	Coast Guard	Black	2,088
53	Coast Guard	Hispanic	2,997
54	Coast Guard	Asian, Hawaiian/Pacific Islander, American Indian/Alaskan Native, Multi-race, Unknown	1,966
Total			1,319,408

Table C-4.
Definition and Control Total of the Dimension DIM4 Used in Raking

DIM4	Service Branch	Region	CTOTAL
11	Army	U.S. - North	38,760
12	Army	U.S. - South, Unknown	235,029
13	Army	U.S. - West	76,642
14	Army	Europe	61,878
15	Army	Asia/Pacific Islands	27,797
16	Army	Other	2,187
21	Navy	U.S. - North	23,349
22	Navy	U.S. - South	158,792
23	Navy	U.S. - West	116,982
24	Navy	Europe	11,705
25	Navy	Asia/Pacific Islands	19,179
26	Navy	Other, Unknown	6,679
31	Marine Corps	U.S. - North	4,010
32	Marine Corps	U.S. - South, Unknown	60,873
33	Marine Corps	U.S. - West	48,732
34	Marine Corps	Europe, Asia/Pacific Islands	14,936
36	Marine Corps	Other	27,711
41	Air Force	U.S. - North	43,307
42	Air Force	U.S. - South, Unknown	144,053
43	Air Force	U.S. - West	98,925
44	Air Force	Europe	35,234
45	Air Force	Asia/Pacific Islands	23,546
46	Air Force	Other	2,682
51	Coast Guard	U.S. - North	8,521
52	Coast Guard	U.S. - South	16,363
53	Coast Guard	U.S. - West	10,746
56	Coast Guard	Other	790
Total			1,319,408

Table C-5.
Definition and Control Total of the Dimension DIM5 Used in Raking

DIM5	Detailed Paygrade	Control Total
11	Enlisted 1, Enlisted 2, Enlisted Unknown	58,948
13	Enlisted 3	201,767
14	Enlisted 4	263,561
15	Enlisted 5	254,759
16	Enlisted 6	179,556
17	Enlisted 7	102,954
18	Enlisted 8	27,418
19	Enlisted 9	10,907
21	Warrant Officer 1	2,052
22	Warrant Officer 2	6,867
23	Warrant Officer 3	4,925
24	Warrant Officer 4	2,436
25	Warrant Officer 5	522
31	Officer 1	18,850
32	Officer 2	29,606
33	Officer 3	69,476
34	Officer 4	44,151
35	Officer 5	28,841
36	Officer 6	11,812
Total		1,319,408

Appendix D.
Variance Estimation

D.1 Variance Estimation

This appendix describes two methodologies that can be used to compute estimates of sampling variability. The first sections include a general review of the two main methods of computing sampling errors or variances of estimates from surveys with complex survey designs, such as the *2005 WEOA*. These methods are linearization (or Taylor series approximation) and replication. The sections also describe software available for computing sampling errors. Standard statistical software packages that assume a simple random sampling design may not properly compute variance estimates from weighted data collected under a design other than simple random sampling. Analyzing weighted *2005 WEOA* data using standard statistical programs will result in accurate point estimates but will not result in accurate variance estimates. While a few features have been described, it is not possible in this setting to compare all features of the three packages.

D.2 Linearization Method to Compute Variances

A widely used method for estimating variances in complex surveys is based on linearization or Taylor series approximation. In this method a linear approximation of a statistic is formed and then substituted into the formula for calculating variance appropriate for the sample design. The linearization method relies on the simplicity associated with estimating the variance for a linear statistic, even with a complex sample design, and is valid in large samples. In this formulation, the variance strata and primary sampling units (PSUs) must be defined. In most complex designs, variance can be estimated by using the variance between PSUs and a replacement estimator (Wolter, 1985). In this formulation, the strata and PSUs must be defined, similar to the variance estimation strata and units discussed earlier. The expression for the variance of a statistic computed from a sample drawn without replacement from stratified ($h = 1$ to H) single stage design is:

$$v(z) = \sum_{h=1}^{H} (1 - f_h) \, n_h \, s_h^2$$

where n_h is the achieved sample in strata h, f_h is the finite population correction (fpc) factor and s_h^2 is computed as

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (z_{hi} - \overline{z}_h)^2}{n_h - 1} \ ,$$

z_{hi} is the appropriate linearized value of the statistic, and $\overline{z}_h = \dfrac{\sum_{i=1}^{n_h} z_{hi}}{n_h}$.

During weighting for the *2005 WEOA*, the variables needed to produce estimates using linearization were also created. The variable TVSTR indicates the variance strata to be used for computing the estimates of variance using the Taylor series method. The variable TVSTR was

created using the sampling strata. Strata with fewer than 30 eligible respondents (with positive final weights) were collapsed with similar strata. Table D-1 presents values for the variable TVSTR for the *2005 WEOA*.

Table D-1.
Assignment of VARSTRAT and Overall Finite Population Factors for Use in WesVar

Variance Strata (TVSTR)	Total Population in Variance Strata (POPTVSTR)	Achieved Sample Size in Variance Strata (SMPTVSTR)	Design Strata
1	103,658	642	1, 7
2	32,155	243	2, 8
3	20,613	86	3, 9
4	2,117	104	4, 10, 34, 40
5	2,690	31	5
6	7,962	104	6, 12, 18, 36, 42, 48
7	2,821	35	11
8	51,860	475	13
9	29,599	376	14
10	10,161	87	15
11	923	85	16, 46
12	2,149	32	17
13	19,293	194	19
14	14,307	186	20
15	3,840	116	21, 51
16	409	50	22, 52
17	737	51	23, 53
18	2,460	42	24, 54
19	45,889	3,985	25
20	7,936	947	26
21	3,105	354	27
22	379	177	28
23	1,695	159	29
24	1,892	214	30
25	28,806	212	31, 37
26	9,342	81	32, 38
27	6,124	326	33, 39
28	968	117	35
29	951	127	41
30	13,501	141	43
31	8,957	108	44
32	2,826	231	45
33	816	147	47
34	3,812	44	49

Table D-1.   (Continued)

Variance Strata (TVSTR)	Total Population in Variance Strata (POPTVSTR)	Achieved Sample Size in Variance Strata (SMPTVSTR)	Design Strata
35	3,785	60	50
36	10,462	1,063	55
37	1,855	217	56
38	694	99	57
39	568	129	58
40	455	50	59
41	72,335	856	60, 66
42	30,221	254	61, 67, 91, 97
43	11,733	143	62, 68
44	5,881	190	63, 69, 93, 99
45	6,375	73	64, 70
46	3,681	43	65, 71, 95, 101
47	66,195	860	72
48	24,631	324	73
49	12,055	160	74
50	2,507	127	75, 81, 105
51	7,127	98	76
52	3,179	51	77, 83, 107, 112
53	21,736	284	78, 108
54	4,971	72	79, 109
55	1,982	59	80, 110
56	1,969	112	82, 111
57	37,881	1,778	84
58	3,538	323	85
59	2,361	171	86
60	186	79	87
61	1,367	128	88
62	1,428	115	89, 117
63	4,209	58	90
64	727	105	92
65	474	38	94
66	3,433	39	96
67	838	110	98
68	588	81	100
69	6,284	108	102
70	3,049	39	103
71	1,438	185	104
72	1,733	266	106
73	3,562	186	113
74	433	46	114
75	265	39	115

Table D-1.   (Continued)

76	244	54	116
77	66,059	311	118, 124
78	9,501	171	119, 125
79	13,923	551	120, 126
80	1,064	108	121, 127
81	2,531	187	122, 128
82	7,074	52	123, 129, 135, 141
83	20,759	162	130
84	6,422	191	131
85	6,177	354	132
86	376	88	133
87	1,056	116	134, 140
88	7,074	74	136
89	3,211	120	137
90	1,301	93	138
91	103	30	139
92	14,008	1,007	142
93	1,219	163	143
94	1,177	157	144
95	102	38	145
96	339	55	146
97	1,248	96	147
98	51,824	875	148
99	12,187	150	149, 179
100	2,664	48	150
101	518	98	151
102	2,539	87	152
103	2,635	46	153, 182
104	27,128	428	154
105	7,587	100	155
106	3,643	49	156
107	185	30	157
108	1,025	30	158
109	2,727	38	159, 187
110	67,681	1,129	160
111	16,198	234	161
112	6,409	101	162
113	459	93	163, 169, 191
114	1,746	73	164, 170, 197
115	6,336	114	165, 171, 193, 198
116	20,638	308	166
117	7,206	95	167, 195
118	1,543	46	168, 196
119	51,593	944	172
120	4,142	243	173

Table D-1.   (Continued)

121	2,250	168	174
122	204	91	175
123	1,266	113	176
124	3,774	214	177, 203
125	7,632	125	178
126	581	72	180
127	301	61	181
128	6,608	85	183
129	2,183	33	184
130	946	104	185
131	245	41	186
132	15,787	254	188
133	4,817	68	189
134	1,788	195	190
135	474	94	192
136	4,395	59	194
137	6,691	117	199
138	541	36	200
139	326	47	201
140	219	41	202
141	10,919	194	204, 208
142	692	65	205, 209
143	1,331	134	206, 210
144	1,816	104	207, 211, 215, 219
145	13,162	301	212
146	1,087	131	213
147	1,168	123	214
148	6,091	179	216
149	379	71	217
150	339	50	218
151	2,337	54	220

D.2.1 Software to Compute Estimates of Variance Using Linearization

SUDAAN© (Research Triangle Institute, 2001) and SAS® (SAS Institute, Inc., 2001) are statistical software packages that can be used to compute estimates of variance for estimates from complex surveys using linearization. These programs include special procedures developed to analyze data from complex surveys. Although the procedures in SAS are more limited than those in SUDAAN, the procedures compute standard errors of the estimates that reflect most features of complex sample designs and nonresponse weighting adjustments. While SUDAAN can also use replication methods, it is most often used for computing variances based on linearization. These programs are also capable of reflecting stratum-by-stratum finite population correction (fpc) factors in the computation of variances. This is particularly important for surveys conducted by DMDC, where some strata are sampled at high rates.

D.3 SUDAAN Procedures

For descriptive statistics, SUDAAN offers three procedures: PROC CROSSTAB for categorical variables, PROC DESCRIPT for continuous variables, and PROC RATIO for ratios of totals. These procedures can be used to compute statistics of interest, such as estimated totals, means, and percentages, along with their corresponding standard errors, design effects, and confidence intervals. SUDAAN can be used to reflect the following in estimating the variance:

- the presence of ineligible members in the frame and the sample (members who become ineligible after the creation of the frame), and

- stratum by stratum finite population correction (fpc) factors

However, SUDAAN cannot reflect variance reduction due to raking. A partial reduction in variance can be reflected if it is assumed that the weights are poststratified to one of the raking dimensions. Using this strategy, the reduction of variance should be evaluated separately for each raking dimension to identify the raking dimension having the most effect on standard errors. SUDAAN reflects the effect of poststratification through the use of POSTVAR and POSTWGT statements, valid in PROC DESCRIPT and PROC RATIO; however, design effects are not computed with this option. Another option in SUDAAN Version 8 is to use replicate weights, in which case the standard errors will be identical to those produced by WesVar™ (Westat, 2000).

Differences of table cell estimates can also be computed in PROC DESCRIPT and PROC RATIO. The statements that control these calculations are CONTRAST, DIFFVAR, and PAIRWISE.

To reflect the effect of the sample design in variance estimation, SUDAAN requires variables that indicate the variance estimation strata and sampled PSUs. The variance estimation strata are generally the original sampling design strata from which the sample was drawn. The sampled PSU corresponds to the individual sampled person. In some design strata the initial sample will be small and will be even further reduced due to nonresponse. Small sample sizes can lead to unstable variance estimates. This problem is limited by collapsing original strata with fewer than 30 respondents.

The variance strata and PSU indicator variables were part of the dataset delivered to DMDC so that estimates and their standard errors could be computed using SUDAAN.

D.4 SAS Procedures

SAS® (SAS Institute, Inc., 2001) has two procedures for analyzing survey data: PROC SURVEYMEANS and PROC SURVEYREG. Both use the Taylor series linearization approach to estimate standard errors. SURVEYMEANS produces estimates of means, proportions, and totals, while SURVEYREG fits linear regression models. No design effects are estimated with either procedure. Estimates of differences or other linear combinations of statistics are not available in SURVEYMEANS.

These procedures are relatively new in SAS and do not contain as many features as some other packages. Accounting for finite population correction factors can be done for variance estimates, but the effect of nonresponse adjustments and raking cannot be accounted for. Accounting for ineligible members in the frame can be done by using the DOMAIN statement, which treats the eligible members as a subpopulation.

D.5 Replication Methods

A second method used to compute estimates of variance is called replication. The basic idea behind replication is to draw subsamples from the full sample, compute the estimate from each of the subsamples, and estimate the variance using the full sample and subsample estimates. The subsamples are called replicates and the estimates from the full and subsamples are called replicate estimates. Rust & Rao (1996) discuss replication methods, show how the units included in the subsamples can be defined using variance strata and units, and describe how these methods can be implemented using weights.

Replicate weights are created to derive a corresponding set of replicate estimates. Each replicate weight is computed using the same estimation steps as the full sample weight, but using only the subsample of cases composing each replicate. The general form of estimates of variance based on replication is:

$$v\left(\hat{\theta}\right) = c\sum_{h=1}^{G}\left(\hat{\theta}_{(k)} - \hat{\theta}\right)^{2},$$

where

$\hat{\theta}$ is the estimate of θ based on the full sample.

$\hat{\theta}_{(k)}$ is the k-th estimate of θ based on the observations included in the k-th replicate.

G is the total number of replicates formed.

c is a constant that depends on the replication method.

$v(\hat{\theta})$ is the estimated variance of $\hat{\theta}$.

An advantage of using replication to estimate variances is the ability to reflect all aspects of weighting: the design, the effect of the nonresponse adjustments, and raking. Through the use of replicates, adjustments made during the weighting process are reflected in the replicate weights applying the same adjustments to each replicate separately.

On the other hand, replication methods have some disadvantages. Replication is computer-intensive, but powerful personal computers have largely eliminated this as an issue. However, it is still possible that for very large datasets the computations will exceed the capacity of the computer or take a long time. Although replication can be used for most estimates, replication techniques are not necessarily appropriate for all sample statistics of interest. Special care is needed when trying to estimate the median, quartiles, or any other quantiles. Another disadvantage is the inclusion of finite population correction (fpc) factors in estimates of variance from sample designs with strata sampled at high rates as in the *2005 WEOA*. In this situation, a variation of a specific replication method needs to be implemented to approximately reflect the finite population correction factors in the computation of variances. The form of this method was specifically developed for weights created for DMDC surveys, where the sample is drawn from deeply stratified designs and some strata are sampled at very high rates.

Replicate weights for computing estimates of variance using replication were created for the *2005 WEOA* data. A special version of the jackknife method was implemented for the *2005 WEOA*. Details of this replication method and its implementation are described in the following sections.

D.5.1 The Jackknife Method

The method of replication used in the *2005 WEOA* is known as the stratified, delete-one jackknife. The general procedure is to form groups of sample members, and then to form replicates or subsamples by deleting one group at a time. The method is also called JKn, and is discussed in Wolter (1985) and Rust (1986).

To implement the method, variables for variance strata (*VARSTRAT*) and variance units (*VARUNIT*) were created. The variance strata are combinations of design strata. The variance units are groups of initial sample members, including eligible and ineligible members, and members with unknown eligibility. Let $\tilde{h}$ be a variance stratum and denote the number of *VARUNIT*s in stratum $\tilde{h}$ by $n_{\tilde{h}}$. Since one *VARUNIT* is omitted at a time in the JKn method, the total number of replicate estimates is:

$$G = \sum_{\tilde{h}=1}^{\tilde{H}} n_{\tilde{h}}$$

where $\tilde{H}$ is the number of variance strata. Note that $\tilde{H}$ may be different from the number of design strata.

Let g denote a particular combination of *VARSTRAT* and *VARUNIT*. Denote the replicate estimate formed by deleting *VARSTRAT-VARUNIT* g by $\hat{\theta}_{(g)}$. Because one *VARUNIT* is omitted at a time for JKn, g can be used to identify the *VARUNIT* itself, the set of sample units (i.e., the replicate) that remains after omitting unit g, and the estimate computed from that replicate set of sample units.

The weights used in calculating $\hat{\theta}_{(g)}$ account for the deletion of g from the sample as follows. Suppose that g identifies a *VARUNIT* in *VARSTRAT* $\tilde{h}$. When *VARSTRAT-VARUNIT* g is omitted, the base weights associated with the other $n_{\tilde{h}} - 1$ variance units in *VARSTRAT* $\tilde{h}$ are multiplied by the factor

$$\frac{n_{\tilde{h}}}{n_{\tilde{h}} - 1}.$$

The base weight for *VARSTRAT-VARUNIT* g is multiplied by 0. The weights on all *VARUNIT*s in all other *VARSTRAT* are unchanged. The two nonresponse adjustment steps and the raking adjustment, described earlier, are then carried through using the sample units in replicate g and their modified base weights. The estimate from replicate g, $\hat{\theta}_{(g)}$, thus reflects all stages of weighting.

The JKn variance estimate for the full sample estimate $\hat{\theta}$ is then:

$$v(\hat{\theta}) = \sum_{g=1}^{G} f_g h_g \left[\hat{\theta}_{(g)} - \hat{\theta}\right]^2$$

where f_g is the finite population correction (fpc) factor associated with the variance stratum containing unit g and $h_g = (n_{\tilde{h}} - 1)/n_{\tilde{h}}$ where $\tilde{h}$ is the *VARSTRAT* that contains unit g. The h_g are referred to as "JKn factors." In forming variance strata, it is important to put design strata having the same or nearly the same fpc together in a variance stratum. This can be done only approximately since the sampling rates vary considerably among the *2005 WEOA* design strata.

D.5.2 *Number of Replicates*

A key step in designing the replicate structure is to determine the number of replicates required. The choice of the number of replicates is based on the desire to obtain an adequate number of degrees of freedom (*df*) to ensure stable estimates of variance, while not having so many as to make the time or cost of computing variance estimates unnecessarily high. At *df* =30, percentiles of the t distribution are near those for the normal distribution; at *df* =60, they are virtually the same as those for the normal. Thus, a rule of thumb is that at least 30 degrees of freedom are needed to obtain relatively stable variance estimates. The stability of a variance estimate for a subgroup is related to the number of VARSTRAT and VARUNITs contributing to the subgroup estimate. Some subgroups, such as white males, are found in many design strata while others, such as members with high disability ratings, are found in few.

Note that having an adequate number of *df* is not a concern in estimates of variances computed by linearization because the estimates will have thousands of degrees of freedom for full sample estimates. Domain estimates will have variances with fewer *df* but probably still enough to insure stability.

D.5.3 Formation of Replicates

The variation of the replication method used to reflect the fpc in estimates of variance was specifically developed by Westat for DMDC surveys. The method has an impact on the way the variance strata (VARSTRAT) are created before the creation of the replicates. The inclusion of the fpc factor is not a straightforward process when replicates are used. As shown in the expression of the variance when JKn replicates are used, the inclusion of the fpc (factor f_g) is only possible at the replicate level. Ideally, the creation of the replicate should be restricted to include the records from a single stratum only, to reflect the effect of the fpc in that specific stratum. At the same time, as described before, to make better estimates at the stratum level, at least 30 replicates per stratum need to be created. Then the total number of replicates to create would be approximated as:

$$\text{Total replicates} \geq 30 * (\text{Number of strata}).$$

The *2005 WEOA* survey has 220 strata, and using the rule above, the required number of replicates needed to fully reflect the fpc in each design stratum would be about 6,600. Such a large number of replicates would be burdensome in practice. To solve this problem, two simplifications were introduced; an overall fpc for groups with similar sampling fractions was used, and design strata were collapsed when the variance strata were created. The fpc for a stratum *h* is:

$$fpc_h = 1 - r_h = 1 - \frac{n_h}{N_h}$$

where:

r_h = the sampling fraction or sampling rate defined as the ratio of the sample size n_h to the total population N_h in stratum *h*.

The pertinent sampling rate here is the achieved rate defined as the number of respondents (not the initial sample size) divided by the population size.

Zones of strata were created such that the design strata within a zone all have approximately the same fpc. The zones were then equated to the *VARSTRAT* for use in WesVar. Table D-2 shows the ranges of stratum sampling rates in each zone and the number of design strata in each.

Table D-2.
Replicate Zones for the 2005 WEOA

Zone	Range of Sampling Rate	Number of Strata	Percentage
1	[0.37, 1.00]	4	1.82
2	[0.18, 0.37)	16	7.27
3	[0.10, 0.18)	40	18.18
4	(0.00, 0.10)	160	72.73
Total		220	100.00

An overall fpc factor was applied to strata within each zone. The overall fpc factor was computed using the minimum sampling rate within the zone. The overall fpc is an approximation of the actual stratum fpc except for the stratum with the minimum sampling rate, where these are the same. In this case, however, the overall fpc is larger than the actual stratum fpc, leading to an overestimation of the variance for estimates for these strata. As a result, this procedure yields somewhat conservative variance estimates. Nevertheless, large improvements are expected in the precision of some domain estimates compared with the case where the fpc is ignored entirely. The fpc for each zone is reported in Table D-3.

Table D-3.
Overall fpc for the Replicate Zones for the 2005 WEOA

Zone	Minimum Sampling Rate	Overall fpc Factor
1	0.3725	0.6275
2	0.1801	0.8199
3	0.1016	0.8984
4	0.0035	0.9965

The design strata can be collapsed (or "folded") into pseudo-strata or variance strata (*VARSTRAT*) to reduce the number of replicates. The number of variance strata and the number of replicates created within each variance stratum affect the number of degrees of freedom of the estimate of variance. As described above, each design stratum should ideally contain at least 30 replicates. The replicate zones were used as variance strata. Table D-4 shows the number of variance strata and number of replicates created within each variance stratum.

Table D-4.
VARSTRAT and VARUNIT for the 2005 WEOA

VARSTRAT	Number of Replicates(VARUNIT)	JKn Factor $\left(h_g\right)$
1	30	0.966667
2	30	0.966667
3	30	0.966667
4	80	0.987500
Total	170	

To assign the value of *VARUNIT*, all records were sorted in the same random order in which they were sampled within *VARSTRAT*. The value of *VARUNIT* is a sequential number starting at 1 and ending at 30 that is assigned to each record. When the sequential number reached the maximum number of *VARUNIT* within *VARSTRAT*, 30, numbering restarts at one. This process was repeated until each member had a *VARUNIT* value. All of the records numbered 1 were assigned to *VARUNIT* 1; all of the records numbered 2 were assigned to *VARUNIT* 2, and so on. The records with *VARUNIT* = 1 were, thus, a subsample of the sample from all design strata assigned to *VARSTRAT* = 1, as were the records in the other *VARUNIT*s. Because the ordering of the sample persons was random, this method effectively divided the sample in each *VARSTRAT* into random groups.

To form the replicates, a series of factors, REPF $\left(\tilde{h}, g\right)$ (replicate factor for *VARUNIT* = g in *VARSTRAT* = $\tilde{h}$), was created with the following values:

$$REPF\left(\tilde{h}, g\right) = \begin{cases} 0 & \text{if the person is in VARSTRAT} = \tilde{h} \text{ and } VARUNIT = g \\ \dfrac{n_{\tilde{h}}}{n_{\tilde{h}} - 1} & \text{if the person is in VARSTRAT} = \tilde{h} \text{ and } VARUNIT \neq g \\ 1 & \text{if the person is in VARSTRAT} \neq \tilde{h} \end{cases}$$

where:

$$n_{\tilde{h}} = \text{the number of } VARUNIT\text{s in } VARSTRAT = \tilde{h}.$$

The replicate base weight is the product of REPF $\left(\tilde{h}, g\right)$ and the full-sample base weight.

The assignment of *VARSTRAT* for the design strata is recorded in Table D-5. The table shows the achieved sampling rate, the actual fpc, and the overall fpc used in each stratum.

D.5.4 Software to Compute Estimates of Variance Using Replication

WesVar™ (Westat, 2000) and SUDAAN are software package that can be used to produce estimates of variance for estimates from complex surveys using replication. While SUDAAN can use replication methods, it is most often used for computing variances based on linearization. Although not fully documented, estimates of variance from most of the replication methods can be implemented in SUDAAN.

D.5.5 WesVar Workbooks

WesVar is a stand-alone computer software program that generates measures of variability (e.g., standard errors, coefficients of variation, and confidence intervals) for estimates using a specified set of replicate weights. Derived statistics, such as differences or ratios, can also be computed in WesVar using the Cell Function feature of tables. WesVar is an interactive application centered on sessions called "workbooks." A workbook is a file linked to a specific WesVar dataset. In a workbook, the user can request descriptive statistics, as well as analyze and create new statistics. The information about the design is incorporated into the replicate weights when the WesVar datafile is created. For descriptive statistics and analysis variables, "requests" are defined within a workbook. Regression requests support both linear and logistic models. Output listings include statistics such as the sum of weights, means, and percentages, along with their corresponding standard errors, design effects, coefficients of variation (CV), and confidence intervals.

Each sample member's record in the datafile has 171 $(G+1)$ weights attached—one for the full sample and 170 (G) replicate sample weights, computed as described above. In WesVar a dataset called a *VAR* file is created that contains an indicator that the JKn method uses to create weights. The *VAR* file also includes the weights themselves, the finite population correction factors, and the h_g factors. When a user requests tabulations or other analyses in WesVar using the *VAR* file, WesVar automatically evaluates variances using the JKn formula.

Table D-5.
Assignment of VARSTRAT and Overall Finite Population Factors (fpc)
for Use in WesVar

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
1	28	0.4670	0.3725	0.5330	0.6275
1	175	0.4461	0.3725	0.5539	0.6275
1	87	0.4247	0.3725	0.5753	0.6275
1	145	0.3725	0.3725	0.6275	0.6275
2	139	0.2913	0.1801	0.7087	0.8199
2	133	0.2340	0.1801	0.7660	0.8199

Table D-5.   (Continued)

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
2	53	0.2316	0.1801	0.7684	0.8199
2	58	0.2271	0.1801	0.7729	0.8199
2	116	0.2213	0.1801	0.7787	0.8199
2	169	0.2113	0.1801	0.7887	0.8199
2	197	0.2059	0.1801	0.7941	0.8199
2	163	0.2051	0.1801	0.7949	0.8199
2	181	0.2027	0.1801	0.7973	0.8199
2	192	0.1983	0.1801	0.8017	0.8199
2	151	0.1892	0.1801	0.8108	0.8199
2	217	0.1873	0.1801	0.8127	0.8199
2	202	0.1872	0.1801	0.8128	0.8199
2	191	0.1842	0.1801	0.8158	0.8199
2	111	0.1808	0.1801	0.8192	0.8199
2	47	0.1801	0.1801	0.8199	0.8199
3	186	0.1673	0.1016	0.8327	0.8984
3	110	0.1641	0.1016	0.8359	0.8984
3	127	0.1632	0.1016	0.8368	0.8984
3	146	0.1622	0.1016	0.8378	0.8984
3	157	0.1622	0.1016	0.8378	0.8984
3	106	0.1535	0.1016	0.8465	0.8984
3	218	0.1475	0.1016	0.8525	0.8984
3	115	0.1472	0.1016	0.8528	0.8984
3	92	0.1444	0.1016	0.8556	0.8984
3	201	0.1442	0.1016	0.8558	0.8984
3	57	0.1427	0.1016	0.8573	0.8984
3	100	0.1378	0.1016	0.8622	0.8984
3	143	0.1337	0.1016	0.8663	0.8984
3	41	0.1335	0.1016	0.8665	0.8984
3	144	0.1334	0.1016	0.8666	0.8984
3	98	0.1313	0.1016	0.8687	0.8984
3	104	0.1287	0.1016	0.8713	0.8984
3	209	0.1277	0.1016	0.8723	0.8984
3	180	0.1239	0.1016	0.8761	0.8984
3	22	0.1226	0.1016	0.8774	0.8984
3	51	0.1221	0.1016	0.8779	0.8984
3	52	0.1209	0.1016	0.8791	0.8984
3	35	0.1209	0.1016	0.8791	0.8984
3	213	0.1205	0.1016	0.8795	0.8984
3	26	0.1193	0.1016	0.8807	0.8984
3	56	0.1170	0.1016	0.8830	0.8984
3	117	0.1167	0.1016	0.8833	0.8984

Table D-5.   (Continued)

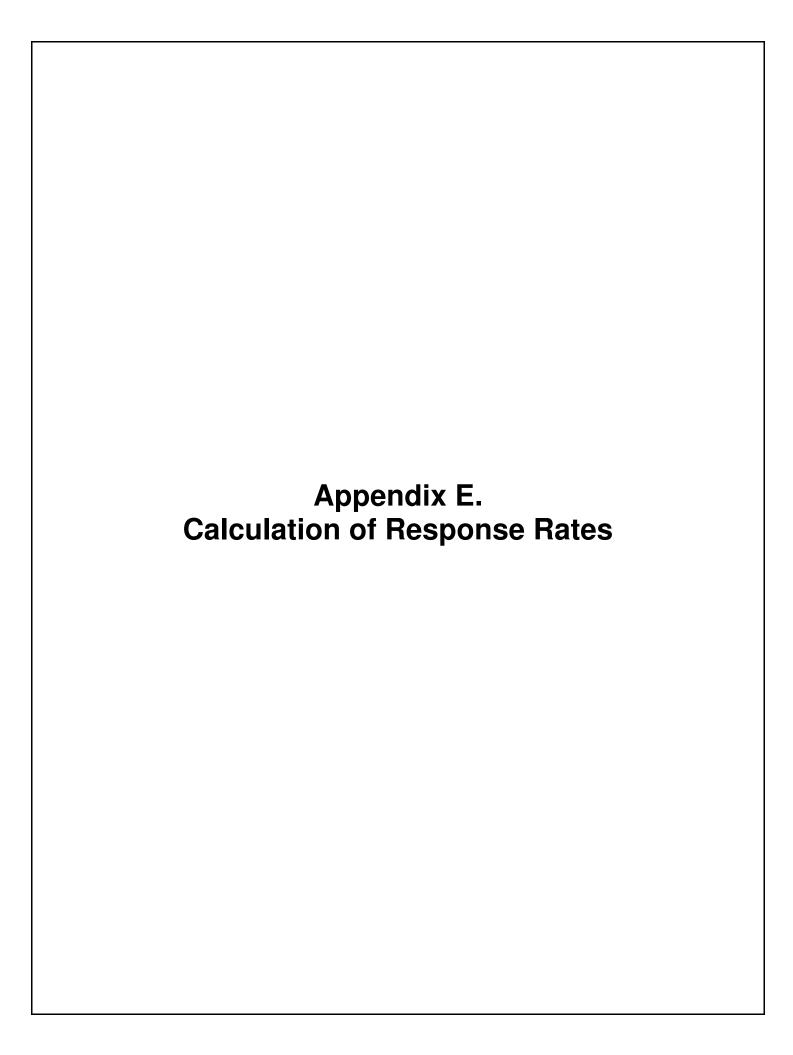| VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
3	27	0.1140	0.1016	0.8860	0.8984
3	140	0.1140	0.1016	0.8860	0.8984
3	30	0.1131	0.1016	0.8869	0.8984
3	185	0.1099	0.1016	0.8901	0.8984
3	59	0.1099	0.1016	0.8901	0.8984
3	190	0.1091	0.1016	0.8909	0.8984
3	134	0.1089	0.1016	0.8911	0.8984
3	196	0.1086	0.1016	0.8914	0.8984
3	46	0.1073	0.1016	0.8927	0.8984
3	114	0.1062	0.1016	0.8938	0.8984
3	210	0.1061	0.1016	0.8939	0.8984
3	214	0.1053	0.1016	0.8947	0.8984
3	55	0.1016	0.1016	0.8984	0.8984
4	206	0.0959	0.0035	0.9041	0.9965
4	29	0.0938	0.0035	0.9062	0.9965
4	88	0.0936	0.0035	0.9064	0.9965
4	85	0.0913	0.0035	0.9087	0.9965
4	176	0.0893	0.0035	0.9107	0.9965
4	16	0.0877	0.0035	0.9123	0.9965
4	25	0.0868	0.0035	0.9132	0.9965
4	128	0.0824	0.0035	0.9176	0.9965
4	45	0.0817	0.0035	0.9183	0.9965
4	94	0.0802	0.0035	0.9198	0.9965
4	121	0.0786	0.0035	0.9214	0.9965
4	147	0.0769	0.0035	0.9231	0.9965
4	89	0.0753	0.0035	0.9247	0.9965
4	174	0.0747	0.0035	0.9253	0.9965
4	219	0.0732	0.0035	0.9268	0.9965
4	86	0.0724	0.0035	0.9276	0.9965
4	142	0.0719	0.0035	0.9281	0.9965
4	138	0.0715	0.0035	0.9285	0.9965
4	122	0.0705	0.0035	0.9295	0.9965
4	81	0.0667	0.0035	0.9333	0.9965
4	200	0.0665	0.0035	0.9335	0.9965
4	205	0.0647	0.0035	0.9353	0.9965
4	105	0.0640	0.0035	0.9360	0.9965
4	34	0.0630	0.0035	0.9370	0.9965
4	215	0.0630	0.0035	0.9370	0.9965
4	40	0.0607	0.0035	0.9393	0.9965
4	33	0.0604	0.0035	0.9396	0.9965
4	173	0.0587	0.0035	0.9413	0.9965

Table D-5.   (Continued)

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
4	132	0.0573	0.0035	0.9427	0.9965
4	177	0.0567	0.0035	0.9433	0.9965
4	203	0.0565	0.0035	0.9435	0.9965
4	10	0.0546	0.0035	0.9454	0.9965
4	113	0.0522	0.0035	0.9478	0.9965
4	75	0.0483	0.0035	0.9517	0.9965
4	84	0.0469	0.0035	0.9531	0.9965
4	99	0.0463	0.0035	0.9537	0.9965
4	39	0.0458	0.0035	0.9542	0.9965
4	207	0.0453	0.0035	0.9547	0.9965
4	211	0.0411	0.0035	0.9589	0.9965
4	126	0.0408	0.0035	0.9592	0.9965
4	120	0.0390	0.0035	0.9610	0.9965
4	69	0.0388	0.0035	0.9612	0.9965
4	137	0.0374	0.0035	0.9626	0.9965
4	4	0.0360	0.0035	0.9640	0.9965
4	152	0.0343	0.0035	0.9657	0.9965
4	164	0.0323	0.0035	0.9677	0.9965
4	131	0.0297	0.0035	0.9703	0.9965
4	216	0.0294	0.0035	0.9706	0.9965
4	158	0.0293	0.0035	0.9707	0.9965
4	170	0.0286	0.0035	0.9714	0.9965
4	93	0.0282	0.0035	0.9718	0.9965
4	63	0.0280	0.0035	0.9720	0.9965
4	220	0.0231	0.0035	0.9769	0.9965
4	107	0.0231	0.0035	0.9769	0.9965
4	212	0.0229	0.0035	0.9771	0.9965
4	125	0.0227	0.0035	0.9773	0.9965
4	208	0.0209	0.0035	0.9791	0.9965
4	83	0.0206	0.0035	0.9794	0.9965
4	171	0.0205	0.0035	0.9795	0.9965
4	95	0.0203	0.0035	0.9797	0.9965
4	112	0.0200	0.0035	0.9800	0.9965
4	82	0.0192	0.0035	0.9808	0.9965
4	97	0.0183	0.0035	0.9817	0.9965
4	172	0.0183	0.0035	0.9817	0.9965
4	165	0.0181	0.0035	0.9819	0.9965
4	150	0.0180	0.0035	0.9820	0.9965
4	23	0.0179	0.0035	0.9821	0.9965
4	153	0.0178	0.0035	0.9822	0.9965
4	199	0.0175	0.0035	0.9825	0.9965

Table D-5.   (Continued)

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
4	24	0.0174	0.0035	0.9826	0.9965
4	102	0.0172	0.0035	0.9828	0.9965
4	148	0.0169	0.0035	0.9831	0.9965
4	160	0.0167	0.0035	0.9833	0.9965
4	198	0.0166	0.0035	0.9834	0.9965
4	178	0.0164	0.0035	0.9836	0.9965
4	193	0.0162	0.0035	0.9838	0.9965
4	188	0.0161	0.0035	0.9839	0.9965
4	119	0.0160	0.0035	0.9840	0.9965
4	141	0.0160	0.0035	0.9840	0.9965
4	50	0.0159	0.0035	0.9841	0.9965
4	154	0.0158	0.0035	0.9842	0.9965
4	162	0.0158	0.0035	0.9842	0.9965
4	54	0.0157	0.0035	0.9843	0.9965
4	18	0.0156	0.0035	0.9844	0.9965
4	48	0.0154	0.0035	0.9846	0.9965
4	184	0.0151	0.0035	0.9849	0.9965
4	80	0.0151	0.0035	0.9849	0.9965
4	79	0.0149	0.0035	0.9851	0.9965
4	166	0.0149	0.0035	0.9851	0.9965
4	17	0.0149	0.0035	0.9851	0.9965
4	182	0.0149	0.0035	0.9851	0.9965
4	68	0.0149	0.0035	0.9851	0.9965
4	159	0.0148	0.0035	0.9852	0.9965
4	161	0.0144	0.0035	0.9856	0.9965
4	189	0.0141	0.0035	0.9859	0.9965
4	90	0.0138	0.0035	0.9862	0.9965
4	76	0.0138	0.0035	0.9862	0.9965
4	167	0.0136	0.0035	0.9864	0.9965
4	156	0.0135	0.0035	0.9865	0.9965
4	194	0.0134	0.0035	0.9866	0.9965
4	204	0.0133	0.0035	0.9867	0.9965
4	74	0.0133	0.0035	0.9867	0.9965
4	155	0.0132	0.0035	0.9868	0.9965
4	73	0.0132	0.0035	0.9868	0.9965
4	78	0.0131	0.0035	0.9869	0.9965
4	20	0.0130	0.0035	0.9870	0.9965
4	66	0.0130	0.0035	0.9870	0.9965
4	72	0.0130	0.0035	0.9870	0.9965
4	108	0.0130	0.0035	0.9870	0.9965
4	183	0.0129	0.0035	0.9871	0.9965

Table D-5.   (Continued)

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
4	103	0.0128	0.0035	0.9872	0.9965
4	70	0.0127	0.0035	0.9873	0.9965
4	14	0.0127	0.0035	0.9873	0.9965
4	11	0.0124	0.0035	0.9876	0.9965
4	149	0.0124	0.0035	0.9876	0.9965
4	77	0.0122	0.0035	0.9878	0.9965
4	44	0.0121	0.0035	0.9879	0.9965
4	179	0.0120	0.0035	0.9880	0.9965
4	195	0.0116	0.0035	0.9884	0.9965
4	129	0.0116	0.0035	0.9884	0.9965
4	49	0.0115	0.0035	0.9885	0.9965
4	5	0.0115	0.0035	0.9885	0.9965
4	71	0.0114	0.0035	0.9886	0.9965
4	96	0.0114	0.0035	0.9886	0.9965
4	60	0.0110	0.0035	0.9890	0.9965
4	67	0.0110	0.0035	0.9890	0.9965
4	65	0.0109	0.0035	0.9891	0.9965
4	187	0.0107	0.0035	0.9893	0.9965
4	168	0.0105	0.0035	0.9895	0.9965
4	136	0.0105	0.0035	0.9895	0.9965
4	43	0.0104	0.0035	0.9896	0.9965
4	64	0.0104	0.0035	0.9896	0.9965
4	109	0.0103	0.0035	0.9897	0.9965
4	19	0.0101	0.0035	0.9899	0.9965
4	12	0.0100	0.0035	0.9900	0.9965
4	13	0.0092	0.0035	0.9908	0.9965
4	91	0.0091	0.0035	0.9909	0.9965
4	21	0.0090	0.0035	0.9910	0.9965
4	38	0.0089	0.0035	0.9911	0.9965
4	42	0.0087	0.0035	0.9913	0.9965
4	135	0.0086	0.0035	0.9914	0.9965
4	62	0.0086	0.0035	0.9914	0.9965
4	15	0.0086	0.0035	0.9914	0.9965
4	32	0.0084	0.0035	0.9916	0.9965
4	36	0.0082	0.0035	0.9918	0.9965
4	8	0.0082	0.0035	0.9918	0.9965
4	37	0.0080	0.0035	0.9920	0.9965
4	130	0.0078	0.0035	0.9922	0.9965
4	101	0.0072	0.0035	0.9928	0.9965
4	6	0.0068	0.0035	0.9932	0.9965
4	31	0.0068	0.0035	0.9932	0.9965

Table D-5.   (Continued)

VARSTRAT	Design Strata	Achieved Sampling Rate	Minimum Sampling Rate Within VARSTRAT	Actual fpc	Overall fpc Within VARSTRAT
4	7	0.0067	0.0035	0.9933	0.9965
4	124	0.0067	0.0035	0.9933	0.9965
4	2	0.0067	0.0035	0.9933	0.9965
4	61	0.0059	0.0035	0.9941	0.9965
4	1	0.0056	0.0035	0.9944	0.9965
4	9	0.0048	0.0035	0.9952	0.9965
4	123	0.0046	0.0035	0.9954	0.9965
4	118	0.0041	0.0035	0.9959	0.9965
4	3	0.0035	0.0035	0.9965	0.9965

Appendix E.
Calculation of Response Rates

Calculation of Response Rates

This appendix describes the formulas used to compute the location, completion, and response rates. The formulas are in accordance with the standards defined by the Council of American Survey Research Organizations (CASRO) and closely follow CASRO's Sample Type II design.

To facilitate computation of the CASRO rates, the variable CAS_ELIG was created to identify the components of *LR*, *CR*, and *RR*. Table E-1 shows the description and distribution of the variable CAS_ELIG.

Table E-1.
Disposition Codes for CASRO Response Rates (CAS_ELIG)

Eligibility Code for CASRO Response Rates (CAS_ELIG)	Description	Sample Cases	Percentage of Sample Cases	Sums of Base Weights	Percentage of Sums of Base Weights
ER	Eligible Respondent-Usable Response	32,299[a]	35.48%	508,818	37.0%
ENR_ACTIVE	Eligible Nonrespondent-Active	93	0.10%	1,236	0.1%
ENR_BLANK	Eligible Nonrespondent-Blank questionnaire returned	24	0.03%	390	0.0%
ENR_NOQCOMP	Eligible Nonrespondent-Incomplete questionnaire returned	3,101	3.41%	45,5670	3.3%
ENR_NORET	Eligible Nonrespondent-Questionnaire not returned-deployed	404	0.44%	5,529	0.4%
IN_SR	Proxy or self Reported ineligible	189	0.21%	2,700	0.2%
UNK_NOLOC	Unknown Eligibility-Nonlocatable or questionnaire not returned	10,710	11.77%	152,209	11.1%
UNK_NORET	Member with unknown eligibility who did not return the questionnaire	40,595	44.60%	603,543	43.8%
IN_FR	Ineligible member identified by the March and July sampling frames	3,609	3.96%	56,880	4.1%
Total		91,024	100.00	1,376,874	100.00

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively

The variable CAS_ELIG was created using the variables ELIG_R, RFLAG_FIN, SCSINEL, SR_ELIG, and COMPFLAG as indicated in Table E-2.

Table E-2.
Creation of the Variable CAS_ELIG

Eligibility Code for CASRO Response Rates (CAS_ELIG)	Weighting Eligibility Code (ELIG_R)	Frame Eligibility (F_ELIG)	Survey Control System Disposition Code (RFLAG_FIN)	SCS Eligibility (SCSINEL)	Self-Reported Eligibility (SR_ELIG)	Complete Questionnaire (COMPFLAG)
ENR_ACTIVE	ENR	1	14, 23	0	1	0, .B
ENR_BLANK	ENR	1	15, 17, 25	0, 13, 14	1	0, .B
ENR_NOQCOMP	ENR	1	1, 7, 8	0, 13, 14	1	0
ENR_NORET	ENR	1	24	14	1	.B
ER	ER	1	1, 7, 8	0, 14	1	1
IN_FR	IN_FR	2	1, 6, 13, 14, 17, 18, 22, 23, 24, 26, 27, 28, 30	0, 2, 9, 12, 14	1, 2	0, 1, .B
IN_SR	IN_PR	1	1, 2, 6, 8, 13 , 18, 19, 22, 25, 30	0, 2, 7, 8, 9, 12	1, 2	0, 1, .B
UNK_NOLOC	UNK	1	27, 28, 29	0	1	.B
UNK_NORET	UNK	1	26	0	1	.B

The expressions for the numbers of located persons, eligible persons, and usable responses in terms of CAS_ELIG are given below. As notational shorthand, CAS_ELIG codes are used to stand for counts of members in the formulas. For example, ER denotes the count of eligible respondents.

The adjusted located sample N_L is defined as the sum of eligible respondents, eligible nonrespondents, and the estimate of members who are assumed to be eligible among the members who did not return the questionnaire. The adjusted located sample, N_L is computed as:

$$N_L = N_{ER} + N_{ENR} + p_E N_{UNK\_NORET}$$

where p_E is the proportion of eligible members observed in the sample computed as:

$$p_E = \frac{N_{ER} + N_{ENR}}{N_{ER} + N_{ENR} + N_{IN\_SR}},$$

and N_{ENR} is the total number of eligible nonrespondent members computed as:

$$N_{ENR} = N_{ENR\_NOQCOMP} + N_{ENR\_BLANK} + N_{ENR\_BLANK}.$$

The adjusted eligible sample N_E is defined as the sum of eligible respondents and the estimate of members who are assumed to be eligible among all members with unknown eligibility. The adjusted eligible sample N_E is computed as:

$$N_E = N_{ER} + N_{ENR} + p_{ER}N_{UNK} \,,$$

where N_{UNK} is the total number of members with unknown eligibility and is computed as:

$$N_{UNK} = N_{UNK\_NORET} + N_{UNK\_NOLOC} \,.$$

The adjusted located count, N_L, and the adjusted eligible count, N_E, can also be expressed by subtracting various counts of ineligible members from the total sample.

The adjusted located count N_L can be computed as:

$$N_L = N - N_{IN} - N_{UNK\_NOLOC} - p_{IN\_SR}N_{UNK} \,,$$

where N is the total number of members computed as $N = N_{ER} + N_{IN} + N_{UNK}$, N_{IN} is the total number of ineligible members observed in the sample computed as $N_{IN} = N_{IN\_FR} + N_{IN\_SR}$, and p_{IN_SR} is the proportion of self-reported or proxy-reported ineligible members observed in the sample, computed as:

$$p_{IN\_SR} = \frac{N_{IN\_SR}}{N_{ER} + N_{IN\_SR}} = 1 - p_{ER} \,.$$

Alternatively, the adjusted eligible count N_E can be computed as:

$$N_E = N - N_{IN} - p_{IN\_SR} \, N_{UNK} \,.$$

Both weighted and unweighted location, completion, and response rates were calculated for the strata used in the sample design as shown in Table E-3.

Table E-3.

Location, Completion, and Response Rates by Design Stratum for the 2005 WEOA

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
001	2,265	1,560	284	68.9	18.2	12.5	68.9	18.2	12.5
002	936	680	85	72.7	12.5	9.1	72.7	12.5	9.1
003	378	260	34	68.8	13.1	9.0	68.8	13.1	9.0
004	274	196	28	71.3	14.3	10.2	71.3	14.3	10.2
005	156	116	31	74.4	26.7	19.9	74.4	26.7	19.9
006	35	24	4	68.6	16.7	11.4	68.6	16.7	11.4
007	1,941	1,561	352	80.4	22.6	18.1	80.4	22.6	18.1
008	1,080	838	156	77.6	18.6	14.5	77.6	18.6	14.5
009	353	270	51	76.5	18.9	14.4	76.5	18.9	14.4
010	259	200	46	77.2	23.0	17.8	77.2	23.0	17.8
011	145	109	35	75.2	32.1	24.1	75.2	32.1	24.1
012	77	58	16	75.3	27.6	20.8	75.3	27.6	20.8
013	1,278	1,149	472	89.9	41.1	36.9	89.9	41.1	36.9
014	1,055	944	375	89.5	39.7	35.6	89.5	39.7	35.6
015	231	210	86	91.0	41.0	37.3	91.0	41.0	37.3
016	167	154	63	92.2	40.9	37.7	92.2	40.9	37.7
017	86	77	31	89.8	40.4	36.2	89.8	40.4	36.2
018	139	124	59	89.4	47.5	42.5	89.4	47.5	42.5
019	335	325	192	97.0	59.1	57.3	97.0	59.1	57.3
020	353	330	185	93.5	56.0	52.4	93.5	56.0	52.4
021	48	46	28	95.8	60.9	58.3	95.8	60.9	58.3
022	60	58	39	96.7	67.2	65.0	96.7	67.2	65.0
023	17	17	10	100.0	58.8	58.8	100.0	58.8	58.8
024	60	60	34	100.0	56.7	56.7	100.0	56.7	56.7
025	6,673	6,372	3,947	95.5	61.9	59.2	95.5	61.9	59.2
026	1,744	1,655	940	94.9	56.8	53.9	94.9	56.8	53.9
027	582	555	346	95.5	62.3	59.5	95.5	62.3	59.5
028	288	279	172	96.6	61.8	59.7	96.6	61.8	59.7
029	288	269	157	93.5	58.4	54.6	93.5	58.4	54.6
030	420	403	210	96.0	52.1	50.0	96.0	52.1	50.0
031	718	619	102	86.2	16.5	14.2	86.2	16.5	14.2
032	301	259	34	86.0	13.1	11.3	86.0	13.1	11.3
033	1,165	1,010	188	86.7	18.6	16.1	86.7	18.6	16.1
034	92	83	16	90.2	19.3	17.4	90.2	19.3	17.4
035	517	450	116	87.1	25.8	22.5	87.1	25.8	22.5
036	24	22	3	91.7	13.6	12.5	91.7	13.6	12.5
037	545	468	110	85.9	23.5	20.2	85.9	23.5	20.2

Table E-3. (Continued)

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
038	255	208	47	81.6	22.6	18.4	81.6	22.6	18.4
039	724	630	137	87.1	21.7	18.9	87.1	21.7	18.9
040	68	61	13	89.7	21.3	19.1	89.7	21.3	19.1
041	514	443	126	86.3	28.4	24.5	86.3	28.4	24.5
042	22	20	4	90.9	20.0	18.2	90.9	20.0	18.2
043	345	319	140	92.5	43.9	40.6	92.5	43.9	40.6
044	307	283	107	92.2	37.8	34.9	92.2	37.8	34.9
045	614	551	231	89.7	41.9	37.6	89.7	41.9	37.6
046	47	43	22	91.5	51.2	46.8	91.5	51.2	46.8
047	336	304	146	90.5	48.0	43.4	90.5	48.0	43.4
048	39	38	17	97.4	44.7	43.6	97.4	44.7	43.6
049	73	70	44	95.9	62.9	60.3	95.9	62.9	60.3
050	109	98	60	89.9	61.2	55.0	89.9	61.2	55.0
051	138	130	87	94.2	67.1	63.2	94.2	67.1	63.2
052	17	17	11	100.0	64.7	64.7	100.0	64.7	64.7
053	62	57	41	91.9	71.9	66.1	91.9	71.9	66.1
054	16	16	8	100.0	50.0	50.0	100.0	50.0	50.0
055	1,796	1,712	1,057	95.3	61.7	58.9	95.3	61.7	58.9
056	444	413	216	93.0	52.3	48.6	93.0	52.3	48.6
057	190	180	98	94.8	54.3	51.5	94.8	54.3	51.5
058	220	211	128	95.9	60.5	58.1	95.9	60.5	58.1
059	95	87	50	91.6	57.5	52.6	91.6	57.5	52.6
060	2,291	1,758	473	76.7	26.9	20.6	76.7	26.9	20.6
061	892	692	95	77.5	13.7	10.6	77.5	13.7	10.6
062	235	195	42	82.8	21.6	17.9	82.8	21.6	17.9
063	562	410	95	73.0	23.2	16.9	73.0	23.2	16.9
064	163	129	36	79.1	27.9	22.1	79.1	27.9	22.1
065	94	75	24	79.4	32.2	25.6	79.4	32.2	25.6
066	1,353	1,105	377	81.6	34.1	27.9	81.6	34.1	27.9
067	582	473	118	81.2	25.0	20.3	81.2	25.0	20.3
068	306	267	100	87.3	37.5	32.7	87.3	37.5	32.7
069	283	243	75	86.2	30.8	26.5	86.2	30.8	26.5
070	123	102	37	82.9	36.3	30.1	82.9	36.3	30.1
071	40	35	11	87.5	31.4	27.5	87.5	31.4	27.5
072	1,902	1,758	859	92.4	48.9	45.2	92.4	48.9	45.2
073	875	782	324	89.4	41.4	37.0	89.4	41.4	37.0
074	362	331	160	91.4	48.3	44.2	91.4	48.3	44.2
075	245	219	104	89.4	47.5	42.4	89.4	47.5	42.4
076	218	201	98	92.2	48.8	45.0	92.2	48.8	45.0
077	54	48	22	88.9	45.8	40.7	88.9	45.8	40.7
078	378	368	261	97.4	71.0	69.1	97.4	71.0	69.1
079	103	103	67	100.0	65.0	65.0	100.0	65.0	65.0

Table E-3.   (Continued)

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
080	37	36	27	97.3	75.0	73.0	97.3	75.0	73.0
081	15	15	10	100.0	66.7	66.7	100.0	66.7	66.7
082	36	36	29	100.0	80.6	80.6	100.0	80.6	80.6
083	20	20	18	100.0	90.0	90.0	100.0	90.0	90.0
084	2,732	2,614	1,770	95.7	67.7	64.8	95.7	67.7	64.8
085	634	596	323	94.0	54.2	50.9	94.0	54.2	50.9
086	305	286	171	93.8	59.8	56.1	93.8	59.8	56.1
087	139	134	79	96.4	59.0	56.8	96.4	59.0	56.8
088	208	202	128	97.1	63.4	61.5	97.1	63.4	61.5
089	169	154	94	91.1	61.0	55.6	91.1	61.0	55.6
090	243	217	58	89.3	26.7	23.9	89.3	26.7	23.9
091	92	69	15	75.0	21.7	16.3	75.0	21.7	16.3
092	391	346	104	88.6	30.1	26.6	88.6	30.1	26.6
093	49	42	8	85.7	19.0	16.3	85.7	19.0	16.3
094	131	117	38	89.3	32.5	29.0	89.3	32.5	29.0
095	13	12	6	92.3	50.0	46.2	92.3	50.0	46.2
096	150	126	39	84.0	31.0	26.0	84.0	31.0	26.0
097	70	61	23	87.1	37.7	32.9	87.1	37.7	32.9
098	366	302	110	82.5	36.4	30.1	82.5	36.4	30.1
099	36	34	10	94.4	29.4	27.8	94.4	29.4	27.8
100	191	171	81	89.5	47.4	42.4	89.5	47.4	42.4
101	7	7	1	100.0	14.3	14.3	100.0	14.3	14.3
102	196	172	108	87.8	62.8	55.1	87.8	62.8	55.1
103	107	93	38	87.2	40.9	35.7	87.2	40.9	35.7
104	390	349	184	89.5	52.7	47.2	89.5	52.7	47.2
105	25	20	13	80.0	65.0	52.0	80.0	65.0	52.0
106	497	451	266	90.7	59.0	53.5	90.7	59.0	53.5
107	9	9	7	100.0	77.8	77.8	100.0	77.8	77.8
108	32	30	22	93.8	73.3	68.8	93.8	73.3	68.8
109	12	11	5	91.7	45.5	41.7	91.7	45.5	41.7
110	49	48	32	98.0	66.7	65.3	98.0	66.7	65.3
111	116	107	83	92.2	77.6	71.6	92.2	77.6	71.6
112	5	5	4	100.0	80.0	80.0	100.0	80.0	80.0
113	280	267	186	95.4	69.7	66.4	95.4	69.7	66.4
114	85	78	46	91.8	59.0	54.1	91.8	59.0	54.1
115	78	72	39	92.3	54.2	50.0	92.3	54.2	50.0
116	79	73	54	92.4	74.0	68.4	92.4	74.0	68.4
117	27	26	21	96.3	80.8	77.8	96.3	80.8	77.8
118	2,037	1,505	206	73.9	13.7	10.1	73.9	13.7	10.1
119	1,251	917	105	73.3	11.4	8.4	73.3	11.4	8.4
120	3,531	2,631	374	74.5	14.2	10.6	74.5	14.2	10.6
121	684	501	61	73.2	12.2	8.9	73.2	12.2	8.9

Table E-3.   (Continued)

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
122	810	602	126	74.4	20.9	15.6	74.4	20.9	15.6
123	235	170	18	72.3	10.6	7.7	72.3	10.6	7.7
124	508	413	103	81.3	24.9	20.3	81.3	24.9	20.3
125	416	323	65	77.6	20.1	15.6	77.6	20.1	15.6
126	1,121	865	175	77.2	20.2	15.6	77.2	20.2	15.6
127	229	183	46	79.8	25.2	20.1	79.8	25.2	20.1
128	270	209	59	77.4	28.2	21.9	77.4	28.2	21.9
129	47	37	10	78.4	26.9	21.1	78.4	26.9	21.1
130	528	469	162	88.8	34.5	30.7	88.8	34.5	30.7
131	671	597	190	89.0	31.8	28.3	89.0	31.8	28.3
132	1,220	1,082	352	88.7	32.5	28.9	88.7	32.5	28.9
133	295	267	88	90.5	33.0	29.8	90.5	33.0	29.8
134	262	228	94	87.0	41.2	35.9	87.0	41.2	35.9
135	59	50	14	84.7	28.0	23.7	84.7	28.0	23.7
136	128	126	72	98.5	57.3	56.4	98.5	57.3	56.4
137	260	241	118	92.8	48.9	45.4	92.8	48.9	45.4
138	206	191	93	92.7	48.7	45.1	92.7	48.7	45.1
139	67	63	30	94.0	47.6	44.8	94.0	47.6	44.8
140	50	47	22	94.0	46.8	44.0	94.0	46.8	44.0
141	19	19	9	100.0	47.4	47.4	100.0	47.4	47.4
142	2,025	1,920	1,002	94.8	52.2	49.5	94.8	52.2	49.5
143	394	369	162	93.7	43.9	41.1	93.7	43.9	41.1
144	342	320	157	93.6	49.1	45.9	93.6	49.1	45.9
145	95	92	38	96.8	41.3	40.0	96.8	41.3	40.0
146	108	102	53	93.8	52.2	48.9	93.8	52.2	48.9
147	232	205	96	88.4	46.8	41.4	88.4	46.8	41.4
148	2,142	1,913	870	89.3	45.5	40.6	89.3	45.5	40.6
149	438	386	127	88.0	32.9	29.0	88.0	32.9	29.0
150	108	100	48	92.6	48.0	44.4	92.6	48.0	44.4
151	277	242	98	87.4	40.5	35.4	87.4	40.5	35.4
152	208	195	87	93.8	44.6	41.8	93.8	44.6	41.8
153	108	92	41	85.2	44.6	38.0	85.2	44.6	38.0
154	1,053	983	426	93.3	43.3	40.4	93.3	43.3	40.4
155	313	288	99	91.8	34.4	31.6	91.8	34.4	31.6
156	132	124	48	94.0	38.8	36.5	94.0	38.8	36.5
157	88	81	30	92.0	37.0	34.1	92.0	37.0	34.1
158	78	69	30	88.5	43.5	38.5	88.5	43.5	38.5
159	94	86	32	91.5	37.2	34.0	91.5	37.2	34.0
160	2,062	2,017	1,126	97.8	55.8	54.6	97.8	55.8	54.6
161	524	506	234	96.6	46.2	44.7	96.6	46.2	44.7
162	179	172	101	96.1	58.7	56.4	96.1	58.7	56.4
163	122	118	64	96.7	54.2	52.5	96.7	54.2	52.5

Table E-3.   (Continued)

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
164	86	83	44	96.5	53.0	51.2	96.5	53.0	51.2
165	132	129	71	97.7	55.0	53.8	97.7	55.0	53.8
166	429	424	306	98.8	72.1	71.3	98.8	72.1	71.3
167	117	117	77	100.0	65.8	65.8	100.0	65.8	65.8
168	20	20	13	100.0	65.0	65.0	100.0	65.0	65.0
169	27	27	15	100.0	55.6	55.6	100.0	55.6	55.6
170	15	14	8	93.3	57.1	53.3	93.3	57.1	53.3
171	24	24	17	100.0	70.8	70.8	100.0	70.8	70.8
172	1,459	1,425	941	97.7	66.1	64.5	97.7	66.1	64.5
173	464	449	242	96.8	53.9	52.1	96.8	53.9	52.1
174	240	233	167	97.1	71.8	69.7	97.1	71.8	69.7
175	154	152	91	98.7	59.9	59.1	98.7	59.9	59.1
176	171	169	113	98.8	66.9	66.1	98.8	66.9	66.1
177	304	297	187	97.7	63.0	61.5	97.7	63.0	61.5
178	366	326	124	89.1	38.1	33.9	89.1	38.1	33.9
179	84	76	22	90.5	28.9	26.2	90.5	28.9	26.2
180	177	155	72	87.6	46.5	40.7	87.6	46.5	40.7
181	159	144	61	90.6	42.4	38.4	90.6	42.4	38.4
182	14	12	5	85.7	41.7	35.7	85.7	41.7	35.7
183	273	245	85	89.7	34.7	31.1	89.7	34.7	31.1
184	91	77	33	84.6	42.9	36.3	84.6	42.9	36.3
185	282	243	104	86.2	42.8	36.9	86.2	42.8	36.9
186	107	98	41	91.6	41.8	38.3	91.6	41.8	38.3
187	26	23	6	88.5	26.1	23.1	88.5	26.1	23.1
188	475	447	253	94.1	56.6	53.2	94.1	56.6	53.2
189	153	141	68	92.2	48.2	44.4	92.2	48.2	44.4
190	409	382	195	93.4	51.0	47.7	93.4	51.0	47.7
191	32	30	14	93.8	46.7	43.8	93.8	46.7	43.8
192	168	160	94	95.2	58.8	56.0	95.2	58.8	56.0
193	45	40	21	88.9	52.5	46.7	88.9	52.5	46.7
194	86	83	58	96.6	70.2	67.8	96.6	70.2	67.8
195	31	28	16	90.3	57.1	51.6	90.3	57.1	51.6
196	56	51	32	89.7	63.3	56.8	89.7	63.3	56.8
197	31	31	21	100.0	67.7	67.7	100.0	67.7	67.7
198	9	9	5	100.0	55.6	55.6	100.0	55.6	55.6
199	197	185	117	93.9	63.2	59.4	93.9	63.2	59.4
200	64	59	36	92.2	61.0	56.3	92.2	61.0	56.2
201	75	69	47	92.0	68.1	62.7	92.0	68.1	62.7
202	78	75	41	96.2	54.7	52.6	96.2	54.7	52.6
203	47	42	27	89.4	64.3	57.4	89.4	64.3	57.4
204	258	233	58	90.2	24.9	22.5	90.2	24.9	22.5
205	153	140	24	91.5	17.1	15.7	91.5	17.1	15.7

Table E-3.   (Continued)

Stratum	Sample Counts			Unweighted			Weighted		
	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
206	235	208	68	88.5	32.7	28.9	88.5	32.7	28.9
207	48	43	15	90.2	34.6	31.3	90.2	34.6	31.3
208	345	327	132	94.9	40.3	38.2	94.9	40.3	38.2
209	116	111	40	95.0	36.2	34.4	95.0	36.2	34.4
210	186	169	66	90.9	39.1	35.5	90.9	39.1	35.5
211	48	44	14	91.7	31.8	29.2	91.7	31.8	29.2
212	535	523	298	97.8	57.0	55.7	97.8	57.0	55.7
213	297	292	129	98.3	44.2	43.5	98.3	44.2	43.5
214	268	262	123	97.8	46.9	45.9	97.8	46.9	45.9
215	90	89	50	98.9	56.2	55.6	98.9	56.2	55.6
216	237	232	178	97.9	76.8	75.2	97.9	76.8	75.2
217	99	95	70	96.0	73.9	70.9	96.0	73.9	70.9
218	74	72	50	97.3	69.4	67.6	97.3	69.4	67.6
219	36	36	24	100.0	66.7	66.7	100.0	66.7	66.7
220	138	121	54	87.7	44.6	39.1	87.7	44.6	39.1

Appendix F.
Software Applications for the Analysis of
the 2005 Workplace and Equal Opportunity
Survey of Active duty Members

Software Applications for the Analysis of the 2005 Workplace and Equal Opportunity Survey of Active Duty Members

Variance estimation procedures have been developed to account for complex sample designs. Using these procedures, the probability of selection of the sample and the use of differential sampling rates for sample subgroups can be appropriately reflected in estimates of sampling error. The two main methods for estimating variances from a complex survey are known as linearization (or Taylor series variance estimation) and replication. Wolter (1985) is a useful reference on the theory and applications of these methods. Shao (1996) is a more recent review paper that compares these methods.

Standard statistical software packages assume use of simple random sampling (SRS) and, therefore, do not properly compute variance estimates from weighted data collected under a design other than SRS. Analyzing the *2005 Workplace and Equal Opportunity Survey of Active duty Members (2005 WEOA)* data with the proper use of the variable RKW0 as the weighting factor in standard statistical programs generally will result in accurate point estimates,[7] but will not result in accurate variance estimates.

This document gives guidance for analyzing the data from the *2005 WEOA* using two software packages (SUDAAN and SAS) that take into account the sampling design of the survey. In general, SUDAAN and SAS produce the same point estimates.[8] The differences that exist between the packages are in the methods used to compute the variances. SUDAAN can use both replication and linearization methods, whereas SAS uses only linearization. Although SAS has a more limited set of statistics available among the two packages, it can still produce most of the statistics typically reported from survey data.

Structure of Data Files

The public release *2005 WEOA* file contains 91,024 records, one for every sampled member. These 91,024 records can be divided into three subgroups which are used for different analytic purposes and may be required by different analytic packages, as shown in Table F-1. The primary analytic subgroup (records with ELIGFLGW = 1) is comprised of the records for eligible respondent members.

---

[7] Differences may occur in point estimates (e.g., means, percentages, and correlations) for different statistical packages as the result of different methods of handling missing data by some procedures.
[8] Since the programs may handle missing values differently, estimates may be different when missing values are present.

Table F-1.
Distribution of Eligibility Flag (ELIGFLGW) in the Public Release 2005 WEOA File

ELIGFLGW	Description	Count	Percentage
1	Eligible respondents	32,299[a]	35.48
2	Self- or proxy-reported ineligible	189	0.21
3	Nonrespondents and frame ineligible	58,536	64.31
Total		91,024	100.00

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively

The second subgroup, (ELIGFLGW = 2), includes the self- or proxy-reported ineligible members not identified as such in the frame. These records were used along with the eligible respondents to develop weights that sum to the population total. Records for the respondents and the self- or proxy-reported ineligible members are used to compute variance estimates based on the linearization method implemented in SUDAAN and SAS. All 32,488 records with ELIGFLGW equal to 1 or 2 should be included in the analytic file when using SUDAAN and SAS.[9] The records for known ineligible members are not used when computing the point estimates, but they are used when computing variances.

The last subgroup, (ELIGFLGW = 3), is composed of nonrespondents and ineligible members identified by the frame. These records are needed only to analyze response rates to the survey and should not be used for any other analyses. If these records are included in the analysis files, SUDAAN and SAS exclude them automatically because they have weights equal to zero.

Records for proxy- or self-reported ineligible members, (ELIGFLGW = 2), should not be excluded when computing variance estimates. As a caution, analysts should not subset the file before passing it to SUDAAN or SAS. Subsetting can result in errors in variance estimates because SUDAAN and SAS do not properly treat the subset as an estimation domain (Valliant, 2002). In this particular situation, SUDAAN and SAS can still estimate variances, though they would be different than the estimates of variance computed using the complete file or all records with nonzero weights.

Analysis of the 2005 WEOA Using SUDAAN

This section describes how to use SUDAAN for the analysis of the 2005 WEOA data and details which options are appropriate to use.

SUDAAN[©] (Software for the Statistical Analysis of Correlated Data) (Research Triangle Institute, 2001) is a statistical package developed by Research Triangle Institute (RTI) to analyze data from complex sample surveys. SUDAAN accounts for the survey design when computing

---

[9] SUDAAN, SAS and WesVar could also process all records in the file. They would simply skip the 58,536 records with zero weights (i.e., RKW0 = 0).

the standard errors of estimates. SUDAAN can use replication methods, but it is most often used for computing variances based on the first-order Taylor series approximation, also known as linearization.

In the last step of weighting for the *2005 WEOA*, the analytical weights are created by raking the nonresponse-adjusted weights to 5 dimensions. SUDAAN cannot reflect the effect of raking on the estimates of variance; therefore, all linearization variance estimates computed using SUDAAN are approximations. In practice, analysts either ignore this effect or make assumptions to partially approximate it. Because SUDAAN can properly account for poststratification, the analyst can assume that the weights were poststratified rather than raked. For the *2005 WEOA*, analysts can use dimension 1 (defined by Service, gender and age groups) as the poststratification cells and proceed as if the weights were poststratified to this dimension. This technique is not recommended when there are large numbers of missing values in one or more of the variables being analyzed. Because SUDAAN creates new poststratified weights when computing estimates with the poststratification option, estimates can be considerably different from the estimates produced when poststratification is ignored. Refer to the SUDAAN manual for details regarding the poststratification option.

Required Variables

The variables that provide information about the sample design in SUDAAN are:

- **Variable TVSTR** (linearization variance strata). The variable TVSTR indicates the variance strata for computing the estimates of variance using the linearization method. The variable TVSTR was created using the sampling strata. Strata with fewer than 30 records with positive final weights were collapsed with similar strata.

- **Variable ELIGFLGW** (final eligibility indicator). The variable ELIGFLGW indicates the final eligibility of the member. Eligible members have ELIGFLGW = 1 while ineligible members have ELIGFLGW = 2. Records with zero final weight have ELIGFLGW = 3.

- **Variable RKW0** (final full sample weight). The variable RKW0 contains the final weight for the full sample. This weight is positive for all the records where ELIGFLGW = 1 or 2.

- **Variable POPTVSTR** (total population in variance strata). The variable POPTVSTR contains the total population for the variance stratum defined by the variable TVSTR. It is required to compute the finite population correction (*fpc*) factors for the estimates of variance.

- **Variable PSTSTR** (final poststratification cell). The variable PSTSTR indicates the final poststratification cell. As mentioned above, in the last weighting adjustment step, final weights were created by raking the nonresponse-adjusted weights to 5 dimensions. Since SUDAAN cannot reflect the variance reduction due to raking, it assumes that the weights are poststratified to one dimension; dimension 1 (DIM 1) in this case. The value of PSTSTR is a sequential number from 1 to 57 corresponding to

the levels of DIM 1. In SUDAAN, the control totals are hard-coded in the program and correspond to totals for cells 1 to 57 in this order.

- **Variables RKW001- RKW170** (final replicate weights). The variables RKW001-RKW170 contain final weights for the 170 replicates created for the *2005 WEOA*. These variables are required when analysts compute variance estimates based on replication methods.

SUDAAN Keywords

The statements and keywords needed to run SUDAAN to compute variance estimates based on linearization are:

- **DESIGN=STRWOR** (required). The *2005 WEOA* implemented a stratified simple random sample design selected without replacement. In some strata the sampling fraction is so large that the *fpc* factor used in the variance estimation formula is nontrivial.

- **NEST TVSTR /MISSUNIT** (required). The keyword NEST lists the variables whose values identify the sampling stages. In this case, the sample was drawn within strata. The option /MISSUNIT instructs SUDAAN to compute the variance contribution of any stratum with only one primary sampling unit (PSU) using the difference of that unit's value and the overall mean value of the population. The file must be sorted by the variable listed in the NEST statement. In the examples that follow this list of statements and keywords, the data sets are already sorted by the variable TVSTR.

- **WEIGHT RKW0** (required). The keyword WEIGHT lists the final weight to be used in the analysis. In this case, the variable for the weight is the final full sample weight RKW0.

- **TOTCNT POPTVSTR** (required if DESIGN=STRWOR). The keyword TOTCNT lists the variable containing the total population count of the strata. In this case, the variable POPTVSTR contains the population totals for the variance stratum TVSTR.

- **SUBPOPN ELIGFLGW=1** (typically required). The keyword SUBPOPN lists the variables and conditions that define the population of interest. The *2005 WEOA* datafile includes ineligible members with a final positive weight. To compute the correct *fpc* factors, the ineligible members should be included in the file. Analyses, however, should be limited to eligible members only (ELIGFLGW = 1). Additional conditions can be included. For example if members in the Army (SRSVC1 = 1) are to be excluded, the statement should be: SUBPOPN ELIGFLGW = 1 & SRSVC1 > 1.

- **POSTVAR PSTSTR** (required but valid only in PROC DESCRIPT and PROC RATIO). The keyword POSTVAR lists the variable that indicates the cells for poststratification. SUDAAN performs an internal poststratification of the final weight, RKW0, using the control totals in the POSTWGT statement. If there are

missing values, SUDAAN computes a new weight different from the final weight given in the WEIGHT statement. This statement cannot be used in PROC CROSSTAB. Also, when the statement POSTVAR is used, the design effect cannot be computed.

- **POSTWGT 136394 86174 63112 51517 28772 9811 3329 25324 14207 9791 7352 4218 1603 689 101160 63659 46726 41814 24008 8572 2852 21559 11323 5799 4678 2893 1643 83652 27732 16192 11125 5673 2372 5825 1903 1788 87291 61290 43809 44121 31739 9272 2406 26919 17004 9626 7332 4726 2212 9537 7528 5210 4365 3852 1976 2586 1366** (Required if POSTVAR is used). This statement follows the statement POSTVAR and lists the control totals for the cells indicated by the variable PSTSTR. These totals correspond to the totals for the raking dimension DIM 1.

The additional statements and keywords needed to run SUDAAN to compute estimates of variance based on replication methods are:

- **DESIGN= JACKKNIFE** (required). The *2005 WEOA* data file includes replicate weights that can be used in SUDAAN. The replication method used to create the weights is a form of the delete-one-group jackknife method. If estimates of variance based on replication methods are computed, the option JACKKNIFE should be used in the design statement.

- **JACKWGTS RKW001- RKW170** (required). The keyword JACKWGTS lists the variable names for the 170 replicate weights created for the *2005 WEOA* data.

- **JACKMULT 30*0.606536156 30*0.792524783 30*0.868540413 80*0.984057823** (required). The keyword JACKMULT lists the 170 replicate factors to be applied to each replicate weight. The factors are computed by multiplying separately the *fpc* factors found in the file FPC_FACT.DAT by the JKn factors found in the file JKN_FACT.DAT for each replicate. Special care is needed so that the order of the factors and the weights are the same in the JACKWGTS and JACKMULT statements and in the files containing the factors.

Estimates Using SUDAAN Based on Linearization

The first example presented in this appendix uses the variable SRSVC1 (Question 2: *In what Service were you on active duty on January 24, 2005?*). The variable SRSVC1 indicates a member's self-reported Service branch. Table F-2 shows the distribution of SRSVC1 for eligible records (ELIGFLGW = 1).

Table F-2.

Distribution of the Variable SRSVC1 (In what Service were you on active duty on January 24, 2005?)

SRSVC1 (In what Service were you on active duty on January 24, 2005?	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
No response	25	0.08	25	0.08
Army	11,816	36.58	11,841	36.66
Navy	7,624	23.6	19,465	60.27
Marine Corps	4,100	12.69	23,565	72.96
Air Force	7,391	22.88	30,956	95.84
Coast Guard	1,343	4.16	32,299[a]	100.00

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

This first example shows how to compute totals, percentages and standard errors for the variable based on linearization, using SUDAAN's PROC CROSSTAB. Figure F-1 shows the SUDAAN code for this example. The procedure CROSSTAB produces weighted frequencies and percentage distributions for univariate and multivariate (single variable or multiple variable) tabulations.

Figure F-1.

SUDAAN Code for PROC CROSSTAB for Marginal Totals, Percentages, and Standard Errors

```
proc crosstab data=WEOA design=strwor deft2;
 weight RKWO;           /* final fs weight */
 nest TVSTR /missunit; /*linearization variance strata */
 totcnt POPTVSTR ;      /*total population in linearization variance strata */
 subpopn ELIGFLGW = 1; /*eligible members only*/
 subgroup SRSVC1 ;
   levels 5;
   tables SRSVC1;
 title 'Figure F-2. Sample PROC CROSSTAB';
 print nsum wsum sewgt deffwgt totper setot defftot  /style=nchs ;
```

The output for this example is shown in Figure F-2. Note that, in this figure, SUDAAN excludes 25 eligible members with missing values of SRSVC1 (i.e., members who did not respond or provided multiple answers to this question) from the analysis. Refer to the SUDAAN manual for details regarding how missing values are handled in the different procedures and statements.

Figure F-2.
Sample PROC CROSSTAB Output of Marginal Totals, Percentages, and Standard Errors

```
                              S U D A A N
             Software for the Statistical Analysis of Correlated Data
          Copyright        Research Triangle Institute     February 2005
                          Release 9.0.1



Number of observations read     :  32488     Weighted count :   1319408
Number of observations skipped :  58536
(WEIGHT variable nonpositive)
Observations in subpopulation  :  32299     Weighted count:   1312934
Denominator degrees of freedom :  32337



Date: 07-12-2005                              Research Triangle Institute
Time: 10:04:42                                    The CROSSTAB Procedure

Variance Estimation Method: Taylor Series (STRWOR)
For Subpopulation: ELIGFLGW = 1
by: In what Service were you on active duty on January 24, 2005?.


-------------------------------------------------------------------------------
In what Service were
  you on active duty                                               DEFF Tot
  on January 24,     Sample  Weighted   SE        DEFF    Tot    SE Tot  Percent
  2005?              Size    Size       Weighted  Weighted Percent Percent #2
-------------------------------------------------------------------------------
Total                32274   1312148.93 3330.58    37.79  100.00    .        .
Army                 11816    439356.99 2158.12     0.43   33.48   0.13     0.24
Navy                  7624    335184.68 1063.32     0.10   25.54   0.09     0.13
Marine Corps          4100    155470.93 2174.15     0.91   11.85   0.15     0.68
Air Force             7391    346191.36  828.29     0.06   26.38   0.08     0.11
Coast Guard           1343     35944.97  292.25     0.09    2.74   0.02     0.06
-------------------------------------------------------------------------------
```

Figure F-3 shows an example of SUDAAN's PROC DESCRIPT[10] as used to compute totals and percentages for SRSVC1. In this example, the statements POSTVAR and POSTWGT are used, and the estimates will partially reflect the reduction in variance due to raking. In this case, the estimates are the same as those produced in the previous example because the weights are poststratified to the control totals in the statement POSTWGT. The standard errors estimated by DESCRIPT are smaller than the CROSSTAB estimates (Figure F-2), because of the effect of poststratification. If poststratification is ignored, PROC DESCRIPT's point estimates and estimates of variance are very close to those from PROC CROSSTAB. The small difference is

---

[10] The procedure DESCRIPT was designed to produce descriptive statistics for continuous variables, but it can also be used for discrete (categorical) variables through combinations of the statements CATLEVEL and VAR and the use of SUDAAN's variable _ONE_.

the result of records with missing values. The output of the code in Figure F-3 is shown in Figure F-4.

Figure F-3.
Example of SUDAAN Code for PROC DESCRIPT

```
proc descript data=WEOA design=strwor;
 weight RKWO;           /* final fs weight */
 nest TVSTR /missunit; /*linearization variance strata */
 totcnt POPTVSTR ;      /*total population in linearization variance strata */
 subpopn ELIGFLGW = 1; /*eligible members only*/
 postvar PSTSTR;
 postwgt
 136394 86174 63112 51517  28772  9811   3329 25324 14207  9791
   7352  4218  1603    689 101160 63659 46726 41814 24008  8572
   2852 21559 11323  5799   4678  2893   1643 83652 27732 16192
  11125  5673  2372  5825   1903   1788 87291 61290 43809 44121
  31739  9272  2406 26919  17004  9626   7332  4726  2212  9537
   7528  5210  4365  3852   1976  2586   1366;
 subgroup SRSVC1 PSTSTR _ONE_;
 levels 5 57 1;
 var SRSVC1 SRSVC1 SRSVC1 SRSVC1 SRSVC1;
 catlevel 1 2 3 4 5;
 table _ONE_;  print nsum wsum  total setotal percent sepercent / style = nchs;
 title 'Figure F-4. Sample PROC DESCRIPT';
 run;
```

Figure F-4.
Sample PROC DESCRIPT Output of Marginal Totals, Percentages, and Standard Errors

```
                          S U D A A N
            Software for the Statistical Analysis of Correlated Data
          Copyright       Research Triangle Institute    February 2005
                          Release 9.0.1



Number of observations read    :  32488    Weighted count :  1319408
Number of observations skipped :  58536
(WEIGHT variable nonpositive)
Observations in subpopulation  :  32299    Weighted count:  1312934
Denominator degrees of freedom :  32337




Date: 07-12-2005                                  Research Triangle Institute
Time: 10:04:54                                      The DESCRIPT Procedure

Variance Estimation Method: Taylor Series (STRWOR)
For Subpopulation: ELIGFLGW = 1
Post-stratified estimates
by: Variable, One.
```

```
--------------------------------------------------------------------------------
Variable            Sample    Weighted
  One               Size      Size        Total      SE Total   Percent   SE Percent
--------------------------------------------------------------------------------
In what Service were
  you on active duty
  on January 24,
  2005?: Army
  Total             32274     1312148.94  439356.76   504.81    33.48     0.03
  1                 32274     1312148.94  439356.76   504.81    33.48     0.03
```

Figure F-4.   (Continued)

```
In what Service were
  you on active duty
  on January 24,
  2005?: Navy
   Total               32274    1312148.94   335184.75    326.51    25.54    0.02
   1                   32274    1312148.94   335184.75    326.51    25.54    0.02
In what Service were
  you on active duty
  on January 24,
  2005?: Marine
  Corps
   Total               32274    1312148.94   155470.97    254.77    11.85    0.02
   1                   32274    1312148.94   155470.97    254.77    11.85    0.02
In what Service were
  you on active duty
  on January 24,
  2005?: Air Force
   Total               32274    1312148.94   346191.52    326.79    26.38    0.02
   1                   32274    1312148.94   346191.52    326.79    26.38    0.02
In what Service were
  you on active duty
  on January 24,
  2005?: Coast Guard
   Total               32274    1312148.94    35944.94    145.53     2.74    0.01
   1                   32274    1312148.94    35944.94    145.53     2.74    0.01
----------------------------------------------------------------------------------
```

Comparing Two Subgroups Using SUDAAN

This example uses the variables SRSVC1 (Question 2: *In what Service were you on active duty on January 24, 2005?*) and EA024 (Question 24: *Overall, how satisfied are you with the military way of life?*). Table F-3 shows the distribution of the variable EA024 for eligible records (ELIGFLGW = 1). SUDAAN will include in the analysis only the members with non-missing values of SRSVC1 and EA024.

Table F-3.
Distribution of the Variable EA024 (Overall, how satisfied are you with the military way of life?)

EA024 (Overall, how satisfied are you with the military way of life?).	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
Multiple response	2	0.01	2	0.01
No response	97	0.30	99	0.31
Very dissatisfied	1,026	3.18	1,125	3.48
Dissatisfied	3,635	11.25	4,760	14.74
Neither satisfied nor dissatisfied	4,538	14.05	9,298	28.79
Satisfied	17,160	53.13	26,458	81.92
Very satisfied	5,841	18.08	32,299[a]	100.00

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

The next example compares the percentages of two subgroups of members who reported being very satisfied with the military way of life (EA024 = 5). The subgroups are members of the Army (SRSVC1 = 1) and members of the Navy (SRSVC1 = 2). For comparing two subgroups within a survey, contrasts can be performed using the PROC DESCRIPT procedure. Figure F-5 shows the SUDAAN code for this example.

Figure F-5.
SUDAAN Code for Comparison of Two Subgroups

```
proc descript data=WEOA design=strwor;
      weight RKWO;                    /* final fs weight */
  nest TVSTR /missunit;              /* linearization variance strata */
  totcnt POPTVSTR;             /*total population in linearization variance strata */
  subpopn ELIGFLGW=1;          /*eligible members only */

  subgroup     EA024          SRSVC1   PSTSTR   _ONE_;
  levels          5      5      57       1 ;
  var EA024 ;
  catlevel  5 ;
  /* the catlevel statement acts as a where statement restricting the analysis to the  fifth
level*/
  /* (in this case) of the variable EA024 (Overall, how satisfied are you with the military way
of life?)*/
  contrast SRSVC1 = (1 -1 0 0 0 ) / name = "Army vs Navy";
  table _ONE_ ;
  print  nsum wsum  total  percent sepercent  t_pct p_pct /style=nchs ;
run ;
```

Figure F-6 shows the output where the estimate of the difference is -3.09, indicating that a higher percentage of Navy members reported being very satisfied with their military life compared to Army members. The *t* statistic for testing if the estimate is different from zero is -4.78 with an associated *p*-value of 0.0000.

Figure F-6.
Sample PROC DESCRIPT Comparison of Two Subgroups

```
                              S U D A A N
             Software for the Statistical Analysis of Correlated Data
            Copyright        Research Triangle Institute    February 2005
                             Release 9.0.1



Number of observations read     :  32488    Weighted count :  1319408
Number of observations skipped  :  58536
(WEIGHT variable nonpositive)
Observations in subpopulation   :  32299    Weighted count:  1312934
Denominator degrees of freedom  :  32337




Date: 07-12-2005                                    Research Triangle Institute
Time: 10:05:06                                         The DESCRIPT Procedure


Variance Estimation Method: Taylor Series (STRWOR)
For Subpopulation: ELIGFLGW = 1
by: Variable, One, Contrast.

for: Variable = Overall, how satisfied are you with the military way of life?:
Very satisfied.


-----------------------------------------------------------------------------------
One                                                                     P-value
   Contrast                                                             T-Test
            Sample   Weighted                         SE Cntrst T-Test   Cont.Pc-
            Size     Size     Cntrst Total Cntrst Pct Pct       Cont.Pct=0 t=0
-----------------------------------------------------------------------------------
Total
   Army vs Navy  19386  771672.03    1560.45     -3.09      0.65      -4.78  0.0000
1
   Army vs Navy  19386  771672.03    1560.45     -3.09      0.65      -4.78  0.0000
-----------------------------------------------------------------------------------
```

Comparing Two Analysis Variables Using SUDAAN

This example compares the responses to two questions within subgroups defined by Service. Comparing two questions within subgroups requires minor data manipulation because SUDAAN does not have an option that easily allows comparison of two analysis variables. If the missing data patterns are the same for the two variables, SAS can be used to create a new

variable containing the differences between the two questions, and, by using the new variable on the VAR statement of the SUDAAN PROC DESCRIPT, produce the *t* statistic in SAS.

The following example uses the variable SRSVC1 (Question 2: *In what Service were you on active duty on January 24, 2005?*), and the variables EA024 (Question 24: *Overall, how satisfied are you with the military way of life?*), and EA016 (Question 16: *Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?*). Table F-4 shows the distribution of the variable EA016 for eligible records (ELIGFLGW = 1).

Table F-4.
Distribution of the Variable EA016 (Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?)

EA016 (Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?).	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
No response	28	0.09	28	0.09
Very unlikely	3,406	10.55	3,434	10.63
Unlikely	4,414	13.67	7,848	24.30
Neither likely nor unlikely	3,625	11.22	11,473	35.52
Likely	10,054	31.13	21,527	66.65
Very likely	10,772	33.35	32,299[a]	100.00

[a] Thirty-one cases were included in the weighting process as eligible respondents but deemed ineligible for reporting purposes. This accounted for the difference between 32,299 and 32,268 usable responses in the statistical methods report and other reports, respectively.

In this example, members who are very likely to stay on active duty (EA016 = 5), are compared to members who reported being very satisfied with the military way of life (EA024 = 5) for 3 subgroups. The subgroups are Army (SRSVC1 = 1), Navy (SRSVC1 = 2), and Air Force (SRSVC1 = 4). Figure F-7 contains the program for this example. In the first part of the program, the SAS code creates auxiliary variables for the analysis variables and computes the difference between these auxiliary variables. In the second part, SUDAAN statements calculate the total, mean, and standard errors of the variable for the difference by subgroups. The means and standard errors of the mean are written to a SAS file to facilitate the analysis. In the last part, SAS code computes the *t* value for the difference in the proportions in the SAS file.

Figure F-7.
Code for Comparison of Two Analysis Variables

```
data temp;
 set WEOA;
 If EAO24 = 5 then a=1; else if EAO24 gt O then a=0;
 if EAO16 = 5 then b=1; else if EAO16 gt O then b=0;
 DIFF=a-b;
 if SRSVC1 = 1 then RSERVICE = 1;
        else

 if SRSVC1= 2 then RSERVICE = 2;
        else
 if SRSVC1= 4 then RSERVICE = 3;
 /*recodes Army to 1, Navy to 2, and Air Force */
 /*to 3 because SUDAAN requires no breaks in code*/
 run;

 proc descript data=temp design=strwor;
 weight RKWO;           /* final fs weight */
 nest TVSTR /missunit; /* linearization variance strata */
 totcnt POPTVSTR;       /*total population in linearization variance strata */

 subpopn ELIGFLGW=1;   /*eligible members only */
subgroup RSERVICE      PSTSTR;
 levels 3       57;
 tables RSERVICE;
 var DIFF ;

 print total setotal mean semean/meanfmt=f10.7 semeanfmt=f10.7 style=nchs;
                      /*output total and mean differences by subgroups*/
 output total setotal mean semean/meanfmt=f10.7 semeanfmt=f10.7
 filename = means filetype = SAS ;
 run;
 data means ;
 set means ;
 mean2 = mean * 100;
 semean2 = semean * 100;
 label mean2 = "% estimate";
 label semean2 = "% stderror";
 tdiff = mean2 / semean2;
 label tdiff = "t value";

 proc print label;
 var RSERVICE total setotal mean2 semean2 tdiff;
 run ;
```

Figure F-8 shows the output for this example. The negative estimates and "large" negative *t* values, indicate that a significantly smaller percentage of members in the Army, the Navy, and the Air Force were very satisfied with military life compared to being very likely to stay on active duty.

Figure F-8.
SUDAAN PROC DESCRIPT and SAS PROC PRINT Listings Showing the Comparison of Two Analysis Variables

```
                              S U D A A N
             Software for the Statistical Analysis of Correlated Data
            Copyright        Research Triangle Institute      February 2005
                              Release 9.0.1


Number of observations read     :  32488    Weighted count :  1319408
Number of observations skipped :  58536
(WEIGHT variable nonpositive)
Observations in subpopulation  :  32299    Weighted count:  1312934
Denominator degrees of freedom :  32337



Date: 07-12-2005                                    Research Triangle Institute
Time: 10:05:45                                         The DESCRIPT Procedure


Variance Estimation Method: Taylor Series (STRWOR)
For Subpopulation: ELIGFLGW = 1
by: Variable, RSERVICE.


-------------------------------------------------------------------------------
Variable
   RSERVICE                   Total        SE Total        Mean      SE Mean
-------------------------------------------------------------------------------
DIFF
   Total               -174903.73         4070.30   -0.1566785    0.0036503
   1                    -56765.97         2774.70   -0.1298054    0.0063521
   2                    -65716.80         2150.32   -0.1967438    0.0064459
   3                    -52420.95         2060.59   -0.1519517    0.0059673

-------------------------------------------------------------------------------
                                                    %           %
Obs    RSERVICE            Total        SE Total  estimate    stderror    t value

 1         0        -174903.73         4070.30   -15.6678     0.36503    -42.9221
 2         1         -56765.97         2774.70   -12.9805     0.63521    -20.4349
 3         2         -65716.80         2150.32   -19.6744     0.64459    -30.5223
 4         3         -52420.95         2060.59   -15.1952     0.59673    -25.4639
```

Estimates Using SUDAAN Based on Replication

As previously mentioned, users of SUDAAN often compute variances based on linearization; however, recent versions of SUDAAN can also compute variances via replication methods. The SUDAAN statements in Figure F-9 produce a table for the variable SRSVC1 similar to the table in Figure F-2.

Figure F-9.
SUDAAN Code for Computing Estimates of Variance Using Replicate Weights

```
proc crosstab data = WEOA  design = JACKKNIFE  ;
weight RKWO;          /* final fs weight */
JACKWGTS RKWO01-RKW170 ;
JACKMULT  30* 0.685113505 30* 0.770764385  30* 0.868725209 80* 0.984057822;
subpopn ELIGFLGW = 1 ; /* eligible members only */
subgroup SRSVC1;
  levels 5;
  tables SRSVC1;
title "Figure F-10 Sample PROC CROSSTAB" ;
print nsum wsum sewgt  totper setot   /style=nchs ;
run ;
```

       The output of this code is shown in Figure F-10. The standard error estimates in Figure F-2 are computed using linearization while the standard error estimates in Figure F-10 are calculated using replication. As shown in these figures, these methods produce the same point estimates (totals and percentages) but the standard errors are slightly different. The standard errors using replication are generally smaller than those computed using linearization because replication methods reflect the reduction of variance due to raking.

Figure F-10.
SUDAAN Output for Computing Estimates of Variance Using Replicate Weights

```
                           S U D A A N
           Software for the Statistical Analysis of Correlated Data
           Copyright        Research Triangle Institute     February 2005
                             Release 9.0.1


Number of observations read    :  32488    Weighted count :  1319408
Number of observations skipped :  58536
(WEIGHT variable nonpositive)
Observations in subpopulation  :  32299    Weighted count:  1312934
Denominator degrees of freedom :    170


Date: 07-12-2005                                  Research Triangle Institute
Time: 11:13:46                                      The CROSSTAB Procedure


Variance Estimation Method: Replicate Weight Jackknife
For Subpopulation: ELIGFLGW = 1
by: In what Service were you on active duty on January 24, 2005?.
--------------------------------------------------------------------------
In what Service were
  you on active duty
  on January 24,        Sample    Weighted     SE           Tot        SE Tot
  2005?                 Size      Size         Weighted     Percent    Percent

--------------------------------------------------------------------------
Total                   32274   1312148.93       834.14     100.00       0.00
Army                    11816    439356.99       563.19      33.48       0.03
Navy                     7624    335184.68       343.74      25.54       0.02
```

```
Marine Corps            4100    155470.93       254.49      11.85       0.02
Air Force               7391    346191.36       341.15      26.38       0.02
Coast Guard             1343     35944.97       142.49       2.74       0.01
-----------------------------------------------------------------------------
```

Comparing Estimates from Different Surveys using SUDAAN

The next example uses a *t* statistic to compare an estimate from one survey with an estimate from an independently selected sample from another survey. The surveys used in this example are the *2005 WEOA* and the *1996 EOS*, which are independent of one another. This example uses the variable EA016 (Question 16 from the *2005 WEOA*: *Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?*) and the variable EQ9628 (Question 18 from the *1996 EOS*: *Suppose that you need to decide whether to remain in the military. Assuming you could remain, how likely is it that you would chose to do so?*), to compare the percentage of military members who reported being likely to remain on active duty in the *2005 WEOA* to those who reported being likely to remain in the military in the *1996 EOS*. Table F-4 shows the distribution of the variable EA016 from the 2005 WEOA and Table F-5 shows the distribution of the variable EQ9628 from the *1996 EOS*.

Table F-5.
Distribution of the 1996 EOS Variable EQ9628 (Suppose that you need to decide whether to remain in the military. Assuming you could remain, how likely is it that you would chose to do so?)

EQ9628 (Suppose that you need to decide whether to remain in the military. Assuming you could remain, how likely is it that you would chose to do so?).	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
Very unlikely	4,911	13.16	4,911	13.16
Unlikely	3,581	9.60	8,492	22.76
Undecided	6,115	16.39	14,607	39.15
Likely	9,673	25.92	24,280	65.07
Very likely	13,033	34.93	37,313	100.00

To compare the proportions, use PROC DESCRIPT as explained in the "Estimates Using SUDAAN Based on Linearization" section above to compute the estimated proportion and standard error of the estimate using the *2005 WEOA* data. The proportion (p_{WEOA}) of WEOA members who reported being likely to remain on active duty is 28.87, with a standard error (se_{WEOA}) of 0.34.

Next, obtain the proportion of members who reported being likely to remain in the military and standard error of the estimate for the *1996 EOS* data. For the EOS data, specify the subpopulation to include only Army, Navy, Air Force, Marine Corps and Coast Guard, excluding

National Guard and Reserve members from the EOS population with the following SUDAAN statement: SUBPOPN ELIGFLGW=1 AND CSERVICE < 6. The proportion (p_{EOS}) of EOS members who reported being likely to remain in the military is 24.09, with a standard error (se_{EOS}) of 0.41.

The difference between the WEOA members and EOS members, computed using the proportions obtained from SUDAAN, is 28.87 – 24.09 = 4.78 percentage points. To compare the proportions p_{WEOA} and p_{EOS}, use the following formula to compute the standard error of the difference:

$$se_{WEOA-EOS} = \sqrt{se_{WEOA}^2 + se_{EOS}^2}$$

and this formula to compute the t statistic for testing the difference:

$$t = \frac{P_{WEOA} - P_{EOS}}{se_{WEOA-EOS}}.$$

In the example above, $se_{WEOA\text{ - }EOS} = \sqrt{(0.34)^2 + (0.41)^2} = 0.53$ percent and $t = \frac{4.78}{0.53} = 8.97$, which shows that a significantly greater percentage of members from the *2005 WEOA* reported being likely to remain on active duty compared to *1996 EOS* members.

Analysis of 2005 WEOA Using SAS

This section describes how to use SAS® (SAS Institute Inc., 1999) to analyze the *2005 WEOA* data.[11] As mentioned before, respondents (ELIGFLGW = 1) and ineligible members (ELIGFLGW = 2) should be kept in the analysis file in order to estimate the variance. The file should include all these cases even if they are not in the subpopulation of interest.

Required Variables

The variables that provide information about the sample design in SAS are:

- **Variable TVSTR** (linearization variance strata). As in SUDAAN, the variable TVSTR indicates the variance strata to be used for computing the estimates of variance using the linearization method.

- **Variable ELIGFLGW** (final eligibility indicator). The variable ELIGFLGW indicates the final eligibility of the member. Eligible members have ELIGFLGW = 1, while ineligible members have ELIGFLGW = 2. Records with zero final weight have ELIGFLGW = 3.

- **Variable RKW0** (final sample weight). The variable RKW0 contains the final weight for the full sample. This weight is positive for all the records where ELIGFLGW = 1 or 2.

---

[11] Examples given in this report were produced using SAS Version Release 9.1.

- **Variable _TOTAL_** (total population in variance strata). SAS requires that the reserved variable name _TOTAL_ be used for the variable that was saved on the data-set as POPTVSTR. This variable contains the population counts for the variance strata (variable TVSTR). It is required to compute the *fpc* factor for the estimates of variance.

Figure F-11 shows the statements[12] available in PROC SURVEYMEANS. The procedure optionally names the input datafile and specifies statistics for the procedure to compute. The VAR statement is required. The VAR statement identifies the variables to be analyzed, whereas the CLASS statement identifies those numeric variables that are to be analyzed as categorical variables. The STRATA statement lists the variables that form the strata in a stratified sample design. The DOMAIN statement lists the variables that define domains for subpopulation analysis. The WEIGHT statement names the sampling weight variable. All statements can appear multiple times except the PROC SURVEYMEANS statement and the WEIGHT statement, which can appear only once.

Figure F-11.
Syntax for SAS PROC SURVEYMEANS

```
PROC SURVEYMEANS < options > < statistic-keywords >;
CLASS variables;
DOMAIN variables < variable*variable variable*variable*variable ... > ;
STRATA variables < / option > ;
VAR variables ;
WEIGHT variable ;
```

In order to take into account finite population correction factors, a file must be created with the reserved SAS variable _TOTAL_. This data set can be either the same file as the one containing the variables to be analyzed, or a new condensed file that is created to speed processing. The statements in Figure F-12 create a working file (MAIN) and a condensed data set (TOTS4FPC) with the stratum population counts. In creating this condensed file, the class statement must specify the stratification variable (i.e., TVSTR) and any variables that are subsequently used in a WHERE statement (e.g., ELIGFLGW).[13]

---

[12] A CLUSTER statement can also be used to specify cluster identification variables in a clustered sample design. A BY statement can be used with PROC SURVEYMEANS to obtain separate analyses for groups defined by the BY variables. Note that using a BY statement provides completely separate analyses of the BY groups unlike the variance estimates when using a DOMAIN statement that takes into account the full variance structure. When a BY statement appears, the procedure expects the input data sets to be sorted in order of the BY variables. The variables are one or more variables in the input data set. If more than one BY statement is specified, the procedure uses only the latest BY statement and ignores any previous ones.

[13] The class statement must also specify any variables in a BY statement to be used in PROC SURVEYMEANS.

Figure F-12.
SAS Code for the Creation of Reduced Data Sets

```
data MAIN ;
  set WEOA
  (keep = ELIGFLGW TVSTR POPTVSTR SRSVC1 RKWO EAO24 EAO16) ;
  *limited variables kept to speed processing ;
  if ELIGFLGW in (1,2) ; *keeps all weighted records ;
  _TOTAL_ = POPTVSTR ; * creates the variable with the SAS required name ;
run ;

proc means data = MAIN noprint;
   var _TOTAL_;
   output out=TOTS4FPC max=;
   class TVSTR ELIGFLGW;
run ;
```

Point Estimates Using SAS

The following statements can be used to compute the proportions of members in each Service using the variable SRSVC1.

Figure F-13.
Sample PROC SURVEYMEANS of Marginal Proportions and Standard Errors Using DOMAIN Statement

```
proc surveymeans data = MAIN total = TOTS4FPC mean stderr;
 strata TVSTR;
 var SRSVC1;
 class SRSVC1;
 domain ELIGFLGW;
 weight RKWO ;
 title 'Figure F-14. Sample SAS PROC SURVEYMEANS';
 run ;
```

The output is shown in Figure 14. The procedure SURVEYMEANS produces proportions and standard errors of proportions, both of which can be converted to percentages by multiplying by 100. The percentages for eligible members match those produced by SUDAAN (Figure F-1). Although not explicitly stated in the output, SAS excludes from the computations all records with missing values of SRSVC1. Missing values can be analyzed as a separate category if the option MISSING is used. Refer to the SAS manual for details of this option and how missing values are handled in this procedure.

Figure F-14.
Sample PROC SURVEYMEANS of Marginal Proportions and Standard Errors Using DOMAIN Statement

```
The SURVEYMEANS Procedure

         Data Summary

Number of Strata                151
Number of Observations        32488
Sum of Weights              1319408

                        Class Level Information

Class
Variable  Label                                                    Levels

SRSVC1    In what Service were you on active duty on January 24, 2005?    6

                        Class Level Information

Class
Variable  Values

SRSVC1    Army Navy Marine Corps Air Force Coast Guard None, you were separated or retired

                                                          Statistics

                                                       Std Error
Variable  Level                           Label    Mean      of Mean
-------------------------------------------------------------------------------
SRSVC1    Army                              '     0.334434     0.001258
          Navy                              '     0.255106     0.000834
          Marine Corps                      '     0.118359     0.001459
          Air Force                         '     0.263468     0.000777
          Coast Guard                       '     0.027361     0.000201
          None, you were separated or retired '   0.001272     0.000255
-------------------------------------------------------------------------------


                        Domain Analysis: Eligible flag

Eligible                                                 Std Error
   flag  Variable  Level                       Label    Mean    of Mean
-------------------------------------------------------------------------------
      1  SRSVC1    Army                          '    0.334838   0.001252
                   Navy                          '    0.255447   0.000837
                   Marine Corps                  '    0.118486   0.001461
                   Air Force                     '    0.263835   0.000780
                   Coast Guard                   '    0.027394   0.000201
                   None, you were separated or retired '     .        .
-------------------------------------------------------------------------------
(removed).
```

Note. Estimates for ineligible members (ELIGFLGW = 2) are removed from the output.

As mentioned previously, the file could be subset to include only eligible members using a WHERE statement in the data step in SAS as shown in the statements in Figure F-15. The code in Figure F-15 computes the proportion of the members in each Service Branch Organization using the variable SRSVC1.[14]

Figure F-15.
Sample PROC SURVEYMEANS of Marginal Proportions and Standard Errors Using WHERE Statement

```
proc surveymeans data = MAIN total = TOTS4FPC mean stderr;
strata TVSTR;
var SRSVC1;
class SRSVC1;
where ELIGFLGW=1;
weight RKWO ;
title 'Figure F-16. Sample SAS PROC SURVEYMEANS';
run;
```

The output is shown in Figure F-16. The percentages match those produced in the previous example, but the standard errors are often smaller than those estimated by SAS when the DOMAIN statement is used. The DOMAIN statement forces SAS to use all weighted cases when estimating the variance structure. The method of using the WHERE statement shown here is not appropriate, however, because it does not take into account the complete probability structure; that is, it is not equivalent to using the SUBPOPN statement in SUDAAN.

Figure F-16.
Sample PROC SURVEYMEANS of Marginal Proportions and Standard Errors Using the WHERE Statement

```
The SURVEYMEANS Procedure

          Data Summary

Number of Strata                151
Number of Observations        32299
Sum of Weights          1312933.89


                     Class Level Information

Class
Variable    Label     Levels    Values
```

---

[14] ELIGFLGW would have to appear on the CLASS statement of the PROC MEAN that creates the file tots4fpc.

```
SRSVC1          '              5    Army Navy Marine Corps Air Force Coast Guard


                                    Statistics

                                                           Std Error
Variable    Level                     Label       Mean      of Mean
-----------------------------------------------------------------------

SRSVC1      Army                        '       0.334838    0.001235
            Navy                        '       0.255447    0.000825
            Marine Corps                '       0.118486    0.001460
            Air Force                   '       0.263774    0.000769
            Coast Guard                 '       0.027386    0.000199
-----------------------------------------------------------------------
```

Comparing Two Subgroups Using SAS

When comparing two subgroups within a survey (e.g., Army vs. Navy), SAS can be used to estimate the difference and variance components, but the *t* test must be calculated separately because it is not possible to request a contrast. This example compares the proportion of Army members (SRSVC1 = 1) with the proportion of Navy members (SRSVC1 = 2) who reported being very satisfied with their military life (EA024 = 5). The statements in Figure F-17 produce the output in Figure F-18.

Figure F-17.
SAS PROC SURVEYMEANS for Comparing Two Subgroups

```
proc surveymeans data = MAIN total= TOTS4FPC mean stderr;
strata TVSTR;
domain SRSVC1*ELIGFLGW;
var EA024;
class EA024;
weight RKWO ;
title 'Figure F-18. Sample SAS PROC SURVEYMEANS';
run ;
```

The difference between the proportions of Army and Navy members who reported being very satisfied with military life is $100*(0.115097 - 0.146011) \approx -3.09$ percentage points. To compare the proportions p_{ARMY} and p_{NAVY}, use the following formula to compute the standard error of the difference:

$$se_{ARMY-NAVY} = \sqrt{se_{ARMY}^2 + se_{NAVY}^2}$$

and the standard formula to compute the *t* statistic for testing the difference:

$$t = \frac{p_{ARMY} - p_{NAVY}}{se_{ARMY-NAVY}}.$$

In the example above, $se_{ARMY\text{-}NAVY} = 100 * \sqrt{(0.004519)^2 + (0.004625)^2} = 0.6466$ percent and $t = \frac{-3.09}{0.6466} = -4.779$, which is equal to the t value produced in the SUDAAN example (any differences are due to rounding).

Figure F-18.
Sample SAS PROC SURVEYMEANS for Comparing Two Subgroups

```
The SURVEYMEANS Procedure

          Data Summary

Number of Strata                 151
Number of Observations         32488
Sum of Weights               1319408

          Class Level Information

Class
Variable     Label       Levels    Values

EA024         '              5     1 2 3 4 5

                 Statistics

                                      Std Error
Variable     Label          Mean       of Mean
-------------------------------------------------
EA024=1                    0.049407    0.002118
EA024=2                    0.140302    0.003040
EA024=3                    0.165413    0.003242
EA024=4       '            0.502554    0.004064
EA024=5                    0.142324    0.002507
-------------------------------------------------


                   Domain Analysis: ELIGFLGW*SRSVC1


Eligible                                                     Std Error
   flag    SRSVC1              Variable    Label      Mean     of Mean
----------------------------------------------------------------------------
    1     Army               EA024=1               0.064766    0.004480
                             EA024=2               0.163200    0.006109
                             EA024=3               0.184350    0.006426
                             EA024=4       '       0.472587    0.007845
                             EA024=5               0.115097    0.004519
          Navy               EA024=1               0.049215    0.003394
                             EA024=2               0.138428    0.005175
                             EA024=3               0.164668    0.005524
                             EA024=4       '       0.501678    0.007199
                             EA024=5               0.146011    0.004625
          Marine Corps       EA024=1               0.070133    0.009195
                             EA024=2               0.161277    0.012253
                             EA024=3               0.186381    0.013077
                             EA024=4       '       0.454848    0.014993
                             EA024=5               0.127360    0.008026
          Air Force          EA024=1               0.024129    0.002001
                             EA024=2               0.106653    0.004019
                             EA024=3               0.135578    0.004459
                             EA024=4       '       0.557249    0.006408
                             EA024=5               0.176391    0.004774
```

```
          Coast Guard          EA024=1                      0.016982            0.004747
                               EA024=2                      0.105454            0.010868
                               EA024=3                      0.140515            0.011875
                               EA024=4        '             0.557718            0.016784
                               EA024=5                      0.179330            0.012070
-------------------------------------------------------------------------------
(removed)
```

Note. Estimates for ineligible members (ELIGFLGW = 2) are removed from the output.

Comparing Two Analysis Variables Using SAS

To compare two questions overall or within subgroups requires manipulating the data to compute the statistical test. If the missing data patterns are the same for the two variables, then SAS can be used to create a new variable containing the differences between the two questions and a *t* statistic can be produced.

To illustrate this, the same variables used in the SUDAAN example that produced Figure 8 are used in this example: EA024 (Question 24: *Overall, how satisfied are you with the military way of life?*) versus EA016 (Question 16: *Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?*). The analysis is not limited to the Army (SRSVC1 = 1), Navy (SRSVC1 = 2), and the Air Force (SRSVC1 = 4) subgroups, as was done for SUDAAN because SAS needs all the weighted cases for variance computation since it does not have a SUBPOPN statement like SUDAAN. The SAS code that computes the differences between the two variables is shown in Figure F-19.

Figure F-19.
Sample PROC SURVEYMEANS Comparison of Two Analysis Variables

```
data main2;
set main;

if EA024 = 5 then a=1; else if EA024 gt 0 then a=0;
if EA016 = 5 then b=1; else if EA016 gt 0 then b=0;
DIFF=a-b;
run;


proc surveymeans data = main2 total = tots4fpc  mean stderr df t;
strata TVSTR;
domain SRSVC1*ELIGFLGW ;
var DIFF ;
weight RKWO;
title 'Figure F-20. Sample SAS PROC SURVEYMEANS';
run;
```

The output is shown in Figure F-20. The estimated percentages exactly match those produced by SUDAAN (Figure 4), with the variances differing only slightly.

Figure F-20.
Sample SAS PROC SURVEYMEANS Comparison of Two Analysis Variables

```
The SURVEYMEANS Procedure

          Data Summary

Number of Strata                  151
Number of Observations          32488
Sum of Weights                1319408


                        Statistics

                               Std Error
Variable        DF          Mean       of Mean    t Value    Pr > |t|
-------------------------------------------------------------------------
DIFF         32039      -0.155374     0.003380     -45.97     <.0001
-------------------------------------------------------------------------



                      Domain Analysis: ELIGFLGW*SRSVC1

Eligible                                           Std Error
flag    SRSVC1          Variable   DF       Mean    of Mean    t Value Pr > |t|
-------------------------------------------------------------------------
   1    Army            DIFF    15661  -0.129805   0.006352     -20.44  <.0001
        Navy            DIFF    11760  -0.196744   0.006446     -30.52  <.0001
        Marine Corps DIFF        4282  -0.133098   0.010390     -12.81  <.0001
        Air Force    DIFF       11731  -0.151952   0.005967     -25.47  <.0001
        Coast Guard  DIFF        1367  -0.209967   0.015940     -13.17  <.0001
(removed)
```

Note. Estimates for ineligible members (ELIGFLGW = 2) are removed from the output.

Combining Multiple Surveys for Analysis

This example provides general guidelines for producing estimates and their standard errors using data combined from two cross-sectional surveys. If research interest is focused upon relatively small populations (e.g., Melanesian- or Polynesian-Americans), the results from any single survey might contain too few respondents to report reliably upon a survey question of interest. In this circumstance, analysts may consider combining the results from multiple surveys to boost the number of respondents and increase the precision of survey estimates. Such combining of survey data raises several methodological issues. These are noted below in the context of an example combining data from the *2005 WEOA* and the *1996 EOS*.

Combining data from different surveys requires, in most cases, restructuring and/or renaming design and analysis variables and adjustment of sample weights for each survey. Analysts should be aware that when combining data sets, it is the data analysts' responsibility to evaluate the comparability of data from the two or more surveys. This example does not require

special design-based analysis because the *2005 WEOA* and *1996 EOS* surveys are comparable in their sampling strata and primary sampling units (PSUs) (i.e., stratified simple random sampling of members). Although these surveys have similar sample designs, there are differences in the weighting methodologies used to adjust the weights for nonresponse and to control totals that prove to have an impact on the variability of estimates.

The combining of surveys raises many issues that analysts must consider. For example, the expected advantage of combining results from two or more independent surveys is an increase in the precision of resulting estimates due to the larger combined sample size. As this example will demonstrate, increases in precision can be more than a simple function of sample size. Another important issue to consider is that, by combining the results of more than one survey, an implicit assumption is made that estimates from the two surveys are invariant with respect to the span of time between the administrations of the surveys. In other words, any differences between the estimates that might be attributable to time are ignored.

As mentioned above, analysts should carefully examine the survey questions that will be combined from the surveys. Not only do the questions need to be equivalent, the response categories should have equivalent definitions as well. The *2005 WEOA* and *1996 EOS* questions used in this example were presented earlier in this appendix in the discussion of comparing estimates from different surveys. In this example, estimates of the percentage of active military members who indicated that they were *likely* or *very likely* to remain in the military by ethnic group were computed. The variables are EA016 from the *2005 WEOA* (Question 16: *Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?*) shown in Table F-4 and the variable EQ9628 from the *1996 EOS* (Question 18: *Suppose that you need to decide whether to remain in the military. Assuming you could remain, how likely is it that you would chose to do so?*) shown in Table F-5.

Another consideration when combining data from surveys (in this case two surveys) is the adjustment of survey weights. Presumably, each survey was weighted to reflect the common population. If combined with no adjustment to the weights, estimates will reflect a population size twice the actual size. There are different methods available for creating a combined weight. In general terms, when combining two surveys the ith combined analysis weight, w_{ci}, for common domains (i.e., eligible domains in both surveys) is

$$w_{ci} = \begin{cases} \alpha w_{1i} & \text{if } w_{1i} \text{ is the weight from Survey 1} \\ (1-\alpha)w_{2i} & \text{if } w_{2i} \text{ is the weight from Survey 2} \end{cases}$$

where $0 \leq \alpha \leq 1$. For domains that were eligible in one survey only, the combined weight corresponds to the analysis weight available from that survey without any adjustment.

There are several ways to select a value for α, for example, computing α so that combined survey estimates have the smallest possible variance. Another common approach is to set α as a function of each survey's sample size. In this example we initially take the latter approach. Since 2005 WEOA and 1996 EOS sample sizes were nearly the same (32,271 and 37,241, respectively), we set $\alpha = 0.50$. In this case, the sum of weights will be estimates of the

total population that corresponds to the population average between 1996 and 2005. Later we discuss the implications of using different values of α.

In order to produce combined estimates, the data from each survey dataset needs to be prepared for combining. Table F-6 shows the common variables used to specify the design in addition to the analysis variables in SUDAAN for the surveys and the equivalent variables in the combined file.

Table F-6.
Variables for Use When Combining 1996 EOS Survey Data with 2005 WEOA Survey Data and the Names of the Combined Variables

Variable Descriptions	2005 WEOA Variables	1996 EOS Variables	Combined Variables
Final Weight	RKW0	ANL_WT	CWEIGHT
Variance Strata	TVSTR	VSTRAT	CSTRATA
Stratum Counts	NVSTRAT	_TOTAL_	CTOTAL
Eligibility Flag	ELIGFLGW	ELIGFLGW	CELIGFLGW
Ethnic affinity	ETH	M_ETH	C_ETH
Analysis Variables	EA016	EQ9628	REMAIN

As shown above, analysts need to rename and recode, as required, design and analysis variables and adjust sample weights using SAS code similar to that shown in Figure F-21. This figure shows how to append the files and create the new variables. First, a new variance strata variable needs to be created (CSTRATA) using the variables VSTRAT and TVSTR so that there is no record with overlapping values of these variables. This code also demonstrates the creation of the combined weight variable (CWEIGHT) by multiplying the analysis weights by 0.50 for members from the Army, Navy, Air Force, Marine Corps and Coast Guard forces (CSERVICE < 6). Note that the weights for the AGR/TAR members National Guard and Reserve that were eligible for the survey (CSERVICE = 6) only in the 1996 EOS are not adjusted. The variable CTOTAL is set to either NVSTRAT or _TOTAL_ depending on the source of the members. The variable for eligibility, CELIGFLGW, is also created reflecting the ineligibility of the members of the National Guard and Reserve from the *1996 EOS* (CELIGFLGW=2). The variable for the likelihood of remaining in the military variable (REMAIN) is created by setting the values of this variable to the values of the variable EQ9628 or EA016, depending on the survey, with the values 1 = *very unlikely* or *unlikely*, 2 = *undecided* and 3 = *likely* or *very likely*. Optional format statements can also be included to ease interpretation of the output.

Figure F-21.
Sample SUDAAN PROC DESCRIPT for Estimates of Combined Surveys

```
data C96_05;
  set EOS96CON (in=_1996) WOE2005 (in=_2005);

   *** Create Combined Data Strata By Appending Strata To Survey Year ***;
  if _1996 then   CStrata=1996*1000+VStrat;
  if _2005 then   CStrata=2005*1000+VStrat;

   *** Create Combined Analysis weight***;
   label CStrata='Combined Strata';
   if _1996 then   do;
             If CService ne 6 then CWeight=ANL_WT/2;
             else CWeight=ANL_WT;
   end;
   if _2005 then  CWeight=RKWO/2;

  *** Create Combined Data Strata Totals Equal to Totals ***;
  label CTotal='Combined Data Strata Totals';
  if _1996 then   CTotal=NVSTRAT;
  if _2005 then   CTotal=_Total_;

  *** Create Combined Data Eligibility Flag by setting AGR/TAR to ineligible for
Combined Dataset ***;
   Label CEligFlgW='Conbined Data Eligibility Flag';
   if _1996 then   do;
             If CService = 6 and EligFlgW=1 then CEligFlgW=2;
             else CEligFlgW=EligFlgW;
   end;
     if _2005 then  CEligFlgW=EligFlgW;

    *** Create and recode Combined Data Analysis Variable ***;
   if _1996 then   do;
```

Table F-7 shows the crosswalk of the detailed ethnicity variable for the two surveys required to create the combined race-ethnicity variable CM_ETH. Note that the categories "other" and "none" are collapsed as one single group because they are not comparable between the *2005 WEOA* and *1996 EOS*. Appropriate code is required in order to create this variable.

Table F-7.
*Crosswalk for the Creation of the Combined Ethnicity Variable CM_ETH using the 2005
WEOA Variable ETH and the 1996 EOS Variable M_ETH*

Ethnic Code Combined Variable CM_ETH		Ethnic Code 2005 Variable ETH	Ethnic Code 1996 Variable M_ETH
0	Unknown	ZZ	0
1	Mexican	AK	1
2	Puerto Rican	AL	2
3	Cuban	AM	3
4	Latin American	AN	4
5	Other Hispanic Descent	AO	5
6	Aleut	AP	6
7	Eskimo	AQ	7
8	Native American Indian	AR	8
9	Chinese	AB	9
10	Japanese	AF	10
11	Korean	AG	11
12	Indian	AA	12
13	Filipino	AC	13
14	Vietnamese	AI	14
15	Other Asian descent	AJ	15
16	Melanesian	AS	16
17	Micronesian	AT	17
18	Polynesian	AU	18
19	Other Pacific Islander Descent	AV	19
20	Guamanian	AD	22
21	Other /None	BG, BH	20, 21

The SUDAAN statements in Figure F-22 calculate the estimates of the percentage of members who were *likely* or *very likely* to remain in the military if they were given the chance to stay or leave for the combined data. The estimates are weighted by the adjusted weight variable, CWEIGHT. For comparison, code similar to this could be written for the individual surveys using the original weights and sample design variables. Table F-8 summarizes the output of the code in Figure 22. The output includes the sample sizes, sum of weights, weighted percentages and standard errors for the ethnic groups for the *2005 WEOA* and the *1996 EOS*, separately and combined.

Figure F-22.
Sample SUDAAN PROC DESCRIPT for Estimates of Combined Surveys

```
proc DESCRIPT data= C96_05 filetype=SAS design=strwor include ;
weight CWeight;
nest CStrata /missunit;
totcnt CTotal;
SUBPOPN CELIGFLGW=1 ;
Class CM_Eth Remain;
var  Remain ;
catlevel 3;
tables CM_Eth ;
print nsum wsum percent sepercent deffpct uppct lowpct /style=nchs wsumfmt=f8.0
totalfmt=f8.0 setotalfmt=f8.0 percentfmt=f8.3 sepercentfmt=f8.3 deffpctfmt=f7.3
uppctfmt=f7.3 lowpctfmt=f7.3;
output nsum wsum percent sepercent / percentfmt=f7.3 sepercentfmt=f5.3
```

Table F-8 shows that the standard errors of the estimates of percentages for ethnic groups from the combined surveys are smaller than the standard errors of the estimates from the *2005 WEOA*. In other words, the estimates from the combined surveys are more precise than the *2005 WEOA* estimates. However, for most of the ethnic groups the standard errors from the combined surveys are larger than the standard errors from the *1996 EOS*. In this case combining the surveys using $\alpha = 0.50$ does not generally improve the precision of the estimates with respect to the *1996 EOS*. These results are unexpected because the surveys being combined have very similar sample sizes and sampling strata.

This finding changes focus from sample size equivalency in setting α to variance minimization strategies. The following discussion considers variance minimization and reveals additional issues analysts may have to address when combining data from multiple surveys.

The mathematical expression for the estimate $\hat{p}_{dc}$, the proportion of the population in domain D, using data from the combined survey S_c can be written as:

$$\hat{p}_{dc} = \frac{\hat{D}_c}{\hat{N}_c} = \frac{\sum_{i \in S_c} w_i' \, \delta_i(d)}{\sum_{i \in S_c} w_i'} \, ,$$

where $\hat{D}_c$ is the estimate of the total population in domain D, $\hat{N}_c$ is the estimate of the total population N, the weight w_i' is the combined weight, and $\delta_i(d)$ is the indicator function defined as:

$$\delta_i(d) = \begin{cases} 1 & i \in D \\ 0 & i \notin D \end{cases} .$$

Table F-8.

Ethnic Group Likelihood Intentions to Remain in the Military for WEOA 2005, EOS 1996, and a Combined Dataset of WEOA 2005 and EOS 1996

Ethnic Group	EOS 1996 Sample Size	Sum of Weights	Percent likely/ very likely to remain in the military	Standard Error	WEOA 2005 Sample Size	Sum of Weights	Percent likely/ very likely to remain in the military	Standard Error	Combined EOS 1996 and WEOA 2005 Sample Size	Sum of Weights	Percent likely/ very likely to remain in the military	Standard Error
Total	37,241	1,310,347	56.16	0.460	32,271	1,312,010	58.63	0.396	69,512	1,311,178	57.40	0.304
Unknown	2,458	39,325	63.45	1.405	2,226	107,497	52.62	1.461	4,684	73,411	55.52	1.146
Mexican	2,725	34,765	52.71	1.159	1,607	45,306	57.01	2.168	4,332	40,036	55.14	1.321
Puerto Rican	1,894	18,557	59.89	1.569	819	21,450	61.20	3.270	2,713	20,003	60.59	1.896
Cuban	196	1,138	50.68	5.907	79	2,087	44.77	9.337	275	1,613	46.85	6.487
Latin American	383	3,559	46.14	3.655	429	11,019	56.38	4.715	812	7,289	53.88	3.671
Other Hispanic Descent	2,889	23,387	47.84	1.413	1,576	37,398	58.06	2.318	4,465	30,393	54.13	1.516
Aleut	28	71	66.23	7.281	7	195	76.72	16.730	35	133	73.94	12.645
Eskimo	47	116	30.73	5.513	18	168	62.05	15.778	65	142	49.22	9.324
N. American Indian	2,460	6,069	54.43	0.865	925	14,709	47.27	3.525	3,385	10,389	49.36	2.560
Chinese	421	1,355	41.25	3.174	143	2,484	36.89	6.791	564	1,920	38.43	4.547
Japanese	544	1,799	57.43	3.295	122	2,559	41.87	6.714	666	2,179	48.29	4.385
Korean	641	2,583	44.44	2.674	317	3,966	62.47	4.600	958	3,275	55.36	3.080
Indian	129	582	43.76	5.738	93	2,621	68.28	7.258	222	1,602	63.83	6.199
Filipino	2,947	23,069	67.59	1.044	1,052	20,572	76.63	2.046	3,999	21,821	71.85	1.117
Vietnamese	234	1,168	28.67	3.774	120	1,818	36.68	7.179	354	1,493	33.55	4.529
Other Asian	1,103	4,919	52.55	2.063	346	5,901	55.73	4.378	1,449	5,410	54.28	2.557
Melanesian	24	147	42.32	11.105	1	7	0.00	0.000	25	77	40.28	10.669
Micronesian	63	366	48.54	8.180	38	1,682	58.61	10.322	101	1,024	56.82	8.635
Polynesian	125	806	65.55	5.743	61	1,299	56.66	9.950	186	1,052	60.06	6.574
Other Pacific Islander	222	983	62.43	4.427	128	2,907	58.56	6.686	350	1,945	59.54	5.116
Guamanian	4	18	6.70	5.264	33	845	60.51	12.055	37	432	59.37	11.822
None or Other	17,704	1,145,565	56.03	0.521	22,131	1,025,517	59.26	0.455	39,835	1,085,541	57.56	0.349

F-35

The expression for $\hat{p}_{dc}$ can be written as a function of the weights (w_{1i} and w_{2i}) from the original surveys S_1 and S_2 and α as:

$$\hat{p}_{dc} = \frac{\sum_{i \in S_1} \alpha \, w_{1i} \, \delta(d) + \sum_{i \in S_2} (1-\alpha) \, w_{2i} \, \delta(d)}{\sum_{i \in S_1} \alpha \, w_{1i} + \sum_{i \in S_2} (1-\alpha) \, w_{2i}} = \frac{\alpha \hat{D}_1 + (1-\alpha)\hat{D}_2}{\alpha \hat{N}_1 + (1-\alpha)\hat{N}_2} = f(\alpha),$$

where $\hat{D}_1$ and $\hat{D}_2$ are the estimates of D and $\hat{N}_1$ and $\hat{N}_2$ are the estimates of N from the surveys S_1 and S_2, respectively. In other words, the estimate $\hat{p}_{dc}$ can be expressed as a weighted average (in terms of α) of the estimates of D and N from the surveys S_1 and S_2. The expression also shows that the estimate $\hat{p}_{dc}$ is a function of α. An optimal value of $\alpha = \alpha_o$ can be obtained so that the estimate $\hat{p}_{dc}$ has the smallest standard error. In other words, α_o would minimize the function

$$se(\hat{p}_{dc}) = \sqrt{v(\hat{p}_{dc})} = \sqrt{var\left(\frac{\alpha \hat{D}_1 + (1-\alpha)\hat{D}_2}{\alpha \hat{N}_1 + (1-\alpha)\hat{N}_2}\right)}.$$

Due to the complexity of the function, a closed form solution for α_{do} is not available. However, using the results for variances of totals as a function of α, the value of α_o can be approximated by

$$\alpha_{do} \cong 1 - \frac{v(\hat{p}_{d1})}{v(\hat{p}_{d1}) + v(\hat{p}_{d2})}.$$

An estimate of α_{do} is

$$\hat{\alpha}_{do} = 1 - \frac{\hat{v}(\hat{p}_{d1})}{\hat{v}(\hat{p}_{d2}) + \hat{v}(\hat{p}_{d2})} = 1 - \frac{v_{d1}}{v_{d1} + v_{d2}},$$

where v_{d1} and v_{d2} are the estimated variances of the percentages $\hat{p}_{d1}$ and $\hat{p}_{d2}$ from surveys S_1 and S_2. The value $\hat{\alpha}_{do}$ is a function of v_{d1} and v_{d2}. If estimates of the surveys have the same precision then $v_{d1} = v_{d2} = v_d$, then the value of $\hat{\alpha}_{do}$ is
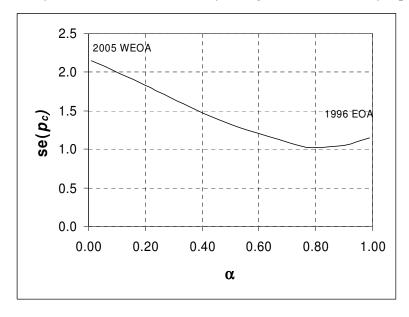
$$\hat{\alpha}_{do} = 1 - \frac{v_d}{v_d + v_d} = 1 - \frac{1}{2} = 0.5.$$

In other words, a value of $\hat{\alpha}_{do} = 0.5$ would produce estimates from the combined surveys with the smallest standard errors when the precision of the estimates from the separate surveys are the same. However, because the values of v_{d1} and v_{d2} for ethnic groups vary within and

across the *2005 WEOA* and *1996 EOS*, a value of $\hat{\alpha}_{do} = 0.5$ is not the optimal value for producing estimates with the minimum standard errors. Furthermore, there is no single value of $\hat{\alpha}_{do}$ that will simultaneously minimize the standard errors of all ethnic group estimates when the data from the *2005 WEOA* and *1996 EOS* are combined. If these observations are generalized, then it is responsibility of the analyst to determine the appropriate value of $\hat{\alpha}_{do}$ for the domain of interest based on the values of the variances of the individual surveys if an increased precision of the combined estimate over the two surveys is required.

As an example of the effect of different values of $\hat{\alpha}_o$ on combined survey estimates, consider the proportion of members of Mexican origin who were *likely* or *very likely* to remain in the military if they were given the chance to stay or leave (third row in Table F-8). The standard error measured as a percentage for this domain for the combined survey is 1.321 ($v_{dc} = 1.745$) for $\alpha_o = 0.5$. This standard error is less than the *2005 WEOA* standard error, 2.168 (v_{d2}=4.700), but larger than the *1996 EOS* standard error, 1.159 (v_{d1}=1.343). An approximation of the optimal value of alpha for this domain is $\alpha_o = 1 - \dfrac{v_{d1}}{v_{d1} + v_{d2}} = 1 - \dfrac{1.343}{1.343 + 4.700} = 0.778$. The standard error for the combined survey using this value for α has a standard error of 1.013 ($v_{dc} = 1.026$), which is smaller than the standard errors for either of the two surveys. The graph in Figure F-23 shows the affect of varying α on the size of the combined variance estimate of the Mexican domain.

Figure F-23.
Standard Error of the Proportion of Members of Mexican Origin who were Likely or Very Likely to Remain in the Military using Combined Surveys (p_c) for Different Values of α

F-37

The expression of the estimated proportion also shows that the choice of α affects the value of the estimate. When surveys are combined the usual assumption is that there is no change in estimates through time. This assumption may or may not be not appropriate in the case for the *1996 EOS* and the *2005 WEOA* with a difference of nine years between administrations. Because a value of $\alpha \neq 0.50$ places more weight on estimates from one survey over the other, the analyst may exacerbate changes due to time in cases when the value of α places more weight on the older survey.

To illustrate the effect of the value of α on the value of the estimate, consider again the percentage of members of Mexican origin who were likely or very likely to remain in the military. The estimate for the *1996 EOS* was 52.71 percent, over four percentage points less than the estimate from the *2005 WEOA* (57.01%). Using $\alpha = 0.50$ puts equal importance on the estimates from each survey when combining, or yields a combined estimate of 55.14 percent, which is close to the simple average of the two estimates. Figure F-24 shows a plot of the point estimate from the combined surveys for different values of α. As α decreases below 0.50 more of importance is given to the *2005 WEOA* estimate producing an estimate closer to 57.01 percent. Conversely, as α increases above 0.50 more importance is given to the *1996 EOS* survey estimate producing an estimate to closer to 52.71%. Using the optimal value of $\alpha_o = 0.778$ the estimate for the combined surveys is 53.87%. This estimate is closest to the *1996 EOS* estimate. If there are time differences between the surveys, it is possible the analyst would report a biased but a more precise estimate.

Figure F-24.
Estimates (p_c) of the Proportion of Members of Mexican Origin Who were Likely or Very Likely to Remain in the Military using Combined Surveys for Different Values of α

Inspection of Table F-8 showed that the standard errors of estimates for ethnic groups from the combined survey are smaller that those from the *2005 WEOA* but they are generally larger than those for the *1996 EOS*. These results are due to the difference in precision of the estimates for these domains between the *1996 EOS* and *2005 WEOA*. As shown above, the optimal value that minimizes the standard error is a function of the variance of the estimates. As a result, a value of $\alpha = 0.50$ may not be appropriate to produce estimates for ethnic groups from combined surveys with the smallest standard errors because the *1996 EOS* estimates are more precise than the *2005 WEOA* estimates. This leads to speculation regarding sources of survey variation.

Although these surveys are very similar in sample design, there are differences in the weighting methodologies used to create the analytical weights (as discussed in detail in the body of this report). To evaluate differences in weighting methodologies, the coefficient of variation (CV) of weights were computed by ethnic groups and are presented in Table F-9. In general, a larger CV for weights implies estimates with greater variability. Although the CV for the total from the *2005 WEOA* is smaller than the corresponding CV from the *1996 EOS,* Table F-9 shows that *2005 WEOA* CVs are larger than *1996 EOS* CVs for most ethnic groups. It is the case that base weight calculations were comparable for the two surveys. However, there were differences in the second and third stages of weighting (nonresponse adjustment and adjustment to control totals). The *1996 EOS* weights were adjusted for nonresponse in one single step and then poststratified to control totals In contrast, the *2005 WEOA* weights were adjusted for nonresponse in two steps and then raked (or simultaneously poststratified) to several control totals. These results suggest the *2005 WEAO* estimates may be less biased but have larger variances due to greater variability in the weights for these domains. However, this observation cannot be verified without a more detailed analysis that compares the estimates and the population from the frames for these domains in from these surveys.

This example of combining survey data from the *2005 WEOA* and the *1996 EOS* highlights several issues analysts must consider and resolve when combining data from multiple surveys. The first is establishing the comparability of questions and their response categories in the two surveys. Next, the sampling and weighting methodologies should be examined to determine whether a degree of similarity exists allowing the combining of survey responses. The examination of weighting methodologies should include an examination of the CVs for estimates of interest for each survey. In preparing the surveys for combining, the most critical decision regards adjustment of the analytical weights for each survey in order to correctly reflect population counts. Two general strategies for setting the value of α have been presented. The first sets α according to the relative sample sizes of the surveys. In this example, the sample sizes were approximately equal so α was set equal to 0.50. (If one survey's sample size were three times that of the other, this strategy would set $\alpha = 0.75$.) This strategy led to the unexpected result where many of the standard errors for the combined survey were greater than those for the *1996 EOS*.

Table F-9.
Coefficient of Variation for the Total and Race-Ethnic Percentage Estimates for the 1996 EOS and 2005 WEOA

	Coefficient of Variation (CV)	
Ethnic group	*1995 EOS*	*2005 WEOA*
Total	160	110
Unknown	116	97
Mexican	79	144
Puerto Rican	102	158
Cuban	139	144
Latin American	111	169
Other Hispanic Descent	123	155
Aleut	43	72
Eskimo	39	91
North American Indian	45	199
Chinese	111	143
Japanese	129	128
Korean	108	146
Indian	100	130
Filipino	86	125
Vietnamese	102	154
Other Asian Descent	106	129
Melanesian	68	-
Micronesian	94	84
Polynesian	96	120
Other Pacific Islander Descent	101	120
Guamanian	100	103
None or Other	108	99

The second strategy presented for setting α was to find a value that produced the smallest standard error. In the case presented, the optimal value was computed to be $\alpha = 0.778$. Use of this strategy raised additional issues. First, α is a weighting parameter that affects the value of the combined estimate. When α was 0.50 the combined estimate was close to the arithmetic mean of the *2005 WEOA* and *1996 EOS* estimates. When set to its optimal value for Mexican respondents, the combined estimate was closer to the estimate reported for the *1996 EOS*. The second issue raised by this strategy is selection of the question or response category to optimize. While $\alpha = 0.778$ is optimal for Mexican respondents, it will not provide the smallest possible standard errors for other ethnic groups.

In summary, there are many decisions and trade-offs analysts must take into account when considering combining data from multiple surveys. As the example above illustrates, the appropriateness of combining survey datasets is dependent upon both the surveys themselves as

well as the manner in which they were administered and processed. Finally, the most appropriate strategy for combining survey datasets may also be a function of what estimates are to be produced from the combined survey dataset.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.